PCT/GB98/00840

/GB 98 / 0 0 8 4 0

16 APR 1998

09/380982

REC'D 16 JUN 1998

WIPO PCT

Europäisches Patentamt | Eur pean Patent Office | Office eur péen des brevets

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

97301917.7

PRIORITY DOCUMENT

Der Präsident des Europäischen Patentamts:
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

M. B. RIJLING

DEN HAAG, DEN
THE HAGUE,
LA HAYE, LE 06/04/98

EPA/EPO/OEB Form 1014 - 02.91

Eur päisches Patentamt
Eur pean Patent Office
Office européen des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.:
Demande n°: 97301917.7

Anmeldetag:
Date of filing:
Date de dépôt: 21/03/97

Anmelder:
Applicant(s):
Demandeur(s):
GeneMark Limited
Potters Bar, Hertfordshire EN6 1AQ
UNITED KINGDOM

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Extraction and utilisation of vntr alleles

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat: State: Pays:	Tag: Date: Date:	Aktenzeichen: File no. Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
C12Q1/68

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

EXTRACTION AND UTILISATION OF VNTR ALLELES.

Glossary of Terms and Abbreviations

AFLP	amplified fragment length polymorphism
allele	one of several possible alternative sequence variations at any one locus
DNA	deoxyribonucleic acid
DNA fingerprint	the display of a set of DNA fragments from a specific DNA sample
individual	a member of any species subject to investigation
heteroduplex	a duplex of two different alleles of any one locus
heterozygygous	alleles at the same locus of each of the paired chromosomes in a diploid cell being different
homoduplex	a duplex of identical alleles of any one locus
homozygous	alleles at the same locus of the paired chromosomes of a diploid cell being identical
locus	a specific position on a chromosome
mis-match	one or more bases in a duplex that fail to form stable hydrogen bonds with apposing bases
NASBA	nucleic acid sequence based amplification
PCR	polymerase chain reaction
RAPD	random-amplified DNA markers
RDA	representational difference analysis
RFLP	restriction fragment length polymorphism
trait	a distinguishing feature or characteristic manifesting itself physically, chemically or biologically

TRAIT	Total Representation of Alleles that are Informative for a Trait
VNTR	variable number tandem repeat, also referred to as simple sequence repeats (encompassing all repeats of two or more nucleotides that may be continuous or interrupted by short non-repetitive sequence, including minisatellites and microsatellites).

Field of the invention

The field of this invention is the detection of polymorphic variation in complex genomes, which is the mainstay of the study of hereditary traits in all organisms. Since polygenic traits far outweigh those that are monogenic, a procedure that allows the isolation in concert of several informative polymorphisms within the complex genomes of multiple individuals would provide an extremely powerful tool for the investigation of hereditary traits.

The invention differs fundamentally from all other techniques that have been previously employed by:

(i) permitting mass generation of VNTRs quickly and easily from DNA

(ii) generating polymorphisms that are both linked and informative for a trait;

(iii) reproducing and preserving the polymorphic allele, as it occurs in the genome;

(iv) negating problems that are features of other polymerase chain reaction based techniques; including miss-priming, reaction contamination and generation of spurious products:

(v) negating the need for investigations to be confined to families of closely-related individuals;

(vi) permitting the analysis of polygenic traits;

(vii) having a sparing requirement for DNA starting material.

The invention therefore represents a major advancement in the ability of workers in the biomedical fields to screen simple or complex genomes, rapidly and with fidelity, for polymorphisms co-segregating with advantageous or deleterious monogenic or polygenic hereditary traits. There is enormous potential for advancement of medicine, veterinary medicine, forensic science, agriculture, animal husbandry and biotechnology, by the generation of polymorphic markers co-segregating with hereditary disease or traits of social or economic importance. The invention will also serve to facilitate mutation analysis for all relevant organisms.

Introduction

DNA is a double stranded linear polymer composed of repetitions of four mononucleotide units. The sequence in which these units are arranged gives rise to a genetic code, referred to as the genome. Although the genomes of all individuals within a species are essentially homologous, subtle variations exist which impart individuality. Locations of the genome at which more than one sequence variation may exist are termed polymorphisms, each variant of that sequence representing an allele. Polymorphisms in gamete-forming germinal cells will be inherited by subsequent generations of progeny. By studying the combination of polymorphisms in the genome of an individual a unique code ('fingerprint') can be assigned and the ancestry of that individual can be determined. Furthermore, a polymorphism found to be linked and co-segregating with a particular genetic trait or hereditary disease may be used as a marker for genetic screening of that trait or disease in other individuals.

The study of advantageous or deleterious hereditary traits in complex genomes has been the subject of considerable interest due to its economical, medical and social implications. The establishment of protocols that allow the comparison of nucleic acid sequences in complex

genomes and the isolation of differences unique to a subset of those sequences is a fundamental requirement of this field of study.

A number of protocols have been used in animals and plants for the comparison of nucleic acid sequences and isolation of differences between those sequences in individuals. These protocols involve restriction fragment length polymorphism (RFLP), random-amplified polymorphic DNA markers (RAPD), amplified fragment length polymorphism (AFLP), representational difference analysis (RDA), and linkage analysis of variable number tandem repeats (VNTR). These protocols detect polymorphisms by assaying subsets of the total DNA sequence variation in a genome. Polymorphisms detected by RFLP, AFLP, and RDA rely on the generation of a fingerprint ladder by gel-electrophoresis which reflects restriction fragment size variation. RAPD polymorphisms result from sequence variation at primer binding sites and differences in length between primer binding sites [1]. Linkage analysis involves the detection of length variation of variable number tandem repeats (VNTRs) and co-segregation of one allele with a trait of interest.

RFLP

RFLP analysis relies on the cleavage of a nucleic acid sequence by restriction endonucleases and separation of the resulting fragments by gel electrophoresis. The fragments are blotted onto a membrane and hybridized to labelled probes to allow detection of fragment length variation [2, 3]. This technique may be of use in the study of a single isolated locus or gene fragment, but where an investigation is not confined to an isolated sequence it is inadequate. Further limitations are that only a small number of the polymorphisms generated may be informative, there is a high demand for DNA starting material, and the method is labour intensive [3].

RAPD

RAPD is a commonly used PCR-based polymorphic marker technique in genomic fingerprinting and diversity studies, particularly for plant species. This technique involves the use of a single 'arbitrary primer' which gives rise to amplification of regions of genome where there is sufficient homology between the sequences of genomic DNA, in the 5' to 3' direction, and that of the arbitrary primer. The amplified products are separated by gel electrophoresis [3]. Subtle variations of this method include arbitrary primed-PCR (AP-PCR) and DNA amplification fingerprinting (DAF). However, the principle of arbitrary priming and amplification of DNA by PCR for difference analysis is common to all. Advantages compared to RFLP are that these methods are more rapid, have a lower demand for DNA, and do not require prior knowledge of sequence. A limitation in common with RFLP is that each analysis can only compare the genomes of two individuals. Although several loci can be evaluated concomitantly by this method, detection of polymorphisms requires observation of variation in band patterns by gel-electrophoresis and is subject to errors of superimposition of different alleles of similar electrophoretic mobility. Many bands may be faint and difficult to interpret, and it is difficult to achieve consistent results in repeat experiments. In common with the majority of PCR techniques, the results are prone to error by subtle changes in reaction conditions, reagent contamination, and the generation of inconsistent banding patterns. This lack of reliability limits the usefulness of such techniques in the 'typing' of individuals [2].

AFLP

AFLP analysis (EP, A, 0534858; Zabeau M *et al.*) involves restriction endonuclease digestion of DNA and ligation of the generated restriction fragments to adaptors. Using primers complementary to the adaptor sequence, the restriction fragments are amplified by PCR, and the

products are separated by gel-electrophoresis, differences in band patterns revealing polymorphisms [2]. Microsatellite-AFLP (WO 96/22388; Kuiper M *et al.*) is a modification of this technique in which two or more restriction enzymes, at least one of which cuts at a simple sequence repeat, are used to cleave DNA into fragments that are ligated to adaptors. The fragments are amplified with primers complementary to the adaptor sequence. In common with RAPD, several loci can be evaluated concomitantly by this method, but detection of polymorphisms requires observation of variation in band patterns by gel-electrophoresis and is subject to errors of superimposition of different alleles of similar electrophoretic mobility. The ability to score bands on an AFLP fingerprint is compromised by generation of large numbers of bands of which some may be very faint and difficult to interpret [6]. Furthermore, the technique is prone to errors that are common to all PCR based techniques, summarised above, and suffers from an inability to analyse multiple complex genomes simultaneously. This is compounded by the generation of bands, by incomplete restriction of the template DNA, that do not reflect true polymorphisms [7]. AFLP and RAPD analyses therefore share many of the same limitations. An additional problem is that AFLPs, rather than being evenly dispersed through out the genome, are reported to be clustered around centromeres [2]. Consequently, this method may not allow the generation of polymorphisms that co-segregate with sequence differences of interest if they are located at a distance from centromeres. This problem is reflected in the reduced rate of polymorphism detection compared to techniques such as linkage analysis [1]. Furthermore, the complexity of the experimental data derived by AFLP becomes exaggerated with increasing complexity of the genome subject to analysis. Consequently, although it has been possible to investigate the genomes of some plant species by AFLP analysis, the relatively complex genomes of higher eukaryotic species may be beyond the useful capacity of this technique.

RDA

RDA involves restriction endonuclease digestion of DNA, ligation of the fragments to adaptors and amplification by PCR. Differences between compared genomes are selected by successive rounds of subtractive hybridization and kinetic enrichment such that regions of difference predominate [5]. This technique is prone to erroneous results through reaction contamination and generation of spurious products. In addition, a fundamental requirement of RDA is the availability of families of closely related individuals, some of which are manifesting the trait of interest. Where RDA is performed on anything other than closely related or highly inbred genomes the multiplicity of differences is too vast for succinct and useful analysis.

In all of the above protocols it is essential that there is a difference in nucleotide sequence at or between primer binding sites or endonuclease restriction sites in order to detect polymorphisms. This highlights the major limitations of these procedures, because in many instances a mutation giving rise to a hereditary trait will not create a sequence difference detectable by variation in primer binding or restriction enzyme digestion. Consequently, a polymorphism linked to a trait of interest will not be identified using these techniques. Furthermore, in contrast to VNTR polymorphisms, primer and restriction site polymorphisms are not hypervariable, and the majority of polymorphisms based on variation in primer binding or restriction enzyme digestion are not informative.

Linkage analysis

Linkage analysis is an indirect molecular genetic strategy that involves the systematic comparison of the inheritance of polymorphic VNTRs with the trait of interest in families in which that trait is present. There are a number of types of VNTR, including minisatellites and

microsatellites, a feature of all of which is the repetition of elements of simple sequences. They are polymorphic by virtue of variation in the number of times each element is repeated, giving rise to alleles with variation in length [2].Since several alternative alleles may exist at any one locus, in contrast to polymorphisms based on variation in primer binding or restriction enzyme digestion, VNTR polymorphic alleles tend to be highly informative. Consequently, where co-segregation of a trait with a particular VNTR allele is demonstrated, the allele may be used as a marker for that trait, or may be used as a vehicle to facilitate identification of the molecular genetic basis of the trait. Microsatellites are ubiquitously distributed throughout all eukaryotic genomes [4]. Consequently, linkage analysis with microsatellites is associated with the highest polymorphism detection rate of the genetic screening methods [1]. Indeed, systematic microsatellite analyses have already been responsible for many advances in the understanding of certain types of common cancer [1]. Linkage analysis therefore has advantages compared to other related methods of difference analysis, the results of which are very reproducible [2]. However, linkage analysis is very time consuming, labour intensive and expensive. Furthermore, since many analyses are performed individually the overall requirement for DNA is extremely high. This is particularly true if a physical map of the genome is unavailable for the selection of informative microsatellites that are evenly distributed throughout the genome. The demonstration of linkage requires the application of elaborate statistical programs and powerful computer software for analysis of the experimental data. This technique is better suited to monogenic defects since the statistical analyses required for multigenic traits are particularly complex. Unfortunately, multifactorial genetic traits are far more prevalent than monogenic defects, making linkage analysis a cumbersome technique for the investigation of the majority of hereditary traits [4].

The characteristics of an ideal protocol for isolation of polymorphisms co-segregating with disease in complex genomes would include:

(i) the ability to isolate simultaneously and with fidelity the polymorphisms from complex genomes of several individuals

(ii) the ability to isolate several polymorphisms simultaneously, permitting the analysis of polygenic traits

(iii) a high detection rate of polymorphisms that co-segregate with sequence differences in all eukaryotic species, including subtle differences such as those resulting from point mutations

(iv) no requirement for large families of closely related individuals to study traits of interest

(v) no requirement for physical maps of the genome or prior knowledge of genomic sequence

(vi) a requirement for sparing quantities of nucleic acid samples for analysis

(vii) simplicity of use without a need for expensive specialist laboratory equipment or computer software

(viii) potential for widespread application throughout the animal and plant kingdoms

(ix) a robust performance with precision, accuracy and fidelity.

None of the techniques that are currently available fulfil the majority of these ideal characteristics. All are compromised by at least one of several limitations including: expense; lack of speed; requirement for large amounts of DNA; low polymorphism detection rate; an inability to detect small sequence variations such as point mutations; a lack of fidelity with high incidence of artefacts and spurious results; inability to analyse several complex genomes concomitantly; an inability to resolve simultaneously polymorphisms at multiple loci; an intrinsic need for closely related genomes for analysis; a need for prior knowledge of sequence; and

complexity of analysis with a need for expensive equipment and computer software. In addition, those techniques that are reliant on large families of closely related individuals are further compromised where there are discrepancies in lineage, so that paternity testing may be an essential preliminary investigation to establish the integrity of each family individual subject to analysis.

The Invention

The invention is a novel method for generating *en masse* the VNTRs from genomic or synthetic DNA, while preserving each allele with its flanking sequence. These alleles may be used to produce a 'fingerprint' by gel electrophoresis, or they may be used as the starting material in protocols for genotyping individuals or protocols for isolation of polymorphic markers that co-segregate with hereditary traits. The latter may be achieved by mis-match discrimination to yield a pool of alleles that are common to all individuals manifesting a particular trait. Further mis-match discrimination of these selected alleles with the alleles of individuals in which the trait is not present, in solution or fixed to an array, allows purification of VNTRs with alleles that are both linked and informative for the particular trait The end products, therefore, are designated a Total Representation of Alleles Informative for a Trait (TRAIT).

In one aspect the invention provides a method of making a mixture of VNTR alleles and their flanking regions of the genomic DNA of one or more members of a species of interest, which method comprises the steps of:

a) dividing genomic DNA of the species of interest into fragments,

b) ligating to each end of each fragment an adaptor having a recessed 3'-end which is blocked to prevent enzymatic chain extension, thereby forming a mixture of adaptor-terminated fragments,

c) using a portion of the mixture of adaptor-terminated fragments as templates with an adaptor primer and a VNTR primer to create a mixture of 5'-flanking VNTR amplimers,

d) using a portion of the mixture of adaptor-terminated fragments as templates with an adaptor primer and a VNTR antisense primer to create a mixture of 3'-flanking VNTR amplimers,

e) and using genomic DNA of the one or more members of the species of interest as template with the mixture of 5'-flanking VNTR amplimers and the mixture of 3'-flanking VNTR amplimers as primers to make the desired mixture of VNTR alleles and their flanking regions.

The species of interest may be any eukaryotic species from the plant and animal kingdoms. Although they do not show repetitive sequences in quite the same way, prokaryotic species are also envisaged. An individual member of a species may be for example a plant or a micro-organism or an animal such as a mammal.

In another aspect the invention provides a portion of genomic DNA of one or more members of a species of interest, said portion consisting essentially of a representative mixture of alleles of a chosen VNTR sequence and their flanking regions.

The term "representative mixture of alleles" does not necessarily imply that all of the possible alleles, or even most of these possible alleles, of a chosen VNTR sequence are present. Whether a particular allele is present or not, e.g. in the mixture generated by the method defined above, may depend on the nature of a restriction enzyme used in step a) and on other factors.

The invention also provides a portion of genomic DNA of a species of interest, said portion consisting essentially of a representative mixture of 3'-flanking regions of a chosen VNTR sequence, each member of the mixture carrying an adaptor at its 3'-end.

The invention also provides a portion of genomic DNA of a

species of interest, said portion consisting essentially of a representative mixture of 5'-flanking regions of a chosen VNTR sequence, each member of the mixture carrying an adaptor at its 5'-end.

The invention also provides a method of treating a mixture of polymorphic alleles, e.g. of a chosen VNTR sequence and their flanking regions, or alternatively a mixture generated in some other way such as AFLP, microsatellite-AFLP or RAPD, the mixture being representative of those which manifest a trait of interest, which method comprises separating and then re-annealing strands of the mixture, and separating and discarding any mis-matches. Preferably the method comprises the additional step of hybridizing the said mixture with a mixture of corresponding polymorphic alleles, e.g. of the chosen VNTR sequence and their flanking regions, or alternatively a mixture generated in some other way such as AFLP, microsatellite-AFLP or RAPD, which are representative of those which do not show the trait of interest, and selecting mis-matches to provide a mixture of polymorphic alleles which are characteristic of the trait of interest.

The invention also provides kits comprising protocols and reagents for performing the methods herein described.

The salient points of the invention may be represented as follows:

(i) reduction in the complexity of the genome by double positive selection of genomic DNA restriction fragments that both ligate to a chosen adaptor and contain a sequence with homology to a chosen primer, employing enrichment of such products by PCR, NASBA or other methods;

(ii) introduction of the selected enriched fragments to a genomic template in such a way that allows recreation of the VNTRs with the flanking sequences within that template, whilst preserving the allele and therefore the informativeness of each locus;

(iii) mis-match discrimination of the generated VNTR alleles to

remove any spurious products of amplification that occur through miss priming events, reaction contamination, and subtle variation in reaction conditions;

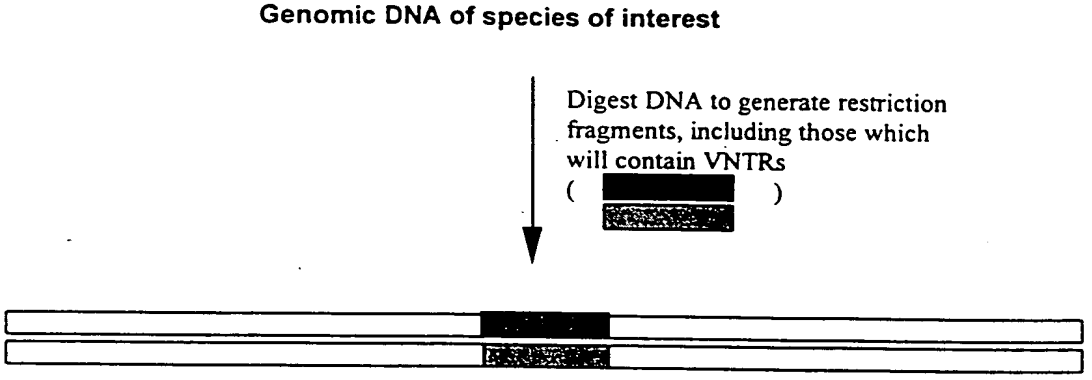
(iv) selection of only those synthesised VNTRs alleles that are common to all individuals manifesting a particular trait. This is achieved by strand dissociation and hybridization, giving rise to mis-match containing heteroduplexes of alleles at any locus that differ among the individuals. These complexes can be rejected by mis-match discrimination. The enriched alleles that are common to all individuals manifesting the trait are sufficiently pure to be used as starting material in other DNA based studies that utilise polymorphic alleles;

(iv) rejection of those alleles common to all individuals manifesting a particular trait that are also common to individuals in which the trait is not present. This is achieved by strand dissociation and hybridization of the VNTR alleles that are common to all individuals manifesting a particular trait of interest with the VNTR alleles of individuals in which the trait is not present followed by a second round of mis-match discrimination. In this case mis-match containing heteroduplexes are selected. These represent polymorphic VNTRs with an informative allele that co-segregates with the particular trait of interest. Amplification of these VNTRs from the DNAs of individuals manifesting the trait of interest yields the informative alleles that may be used as DNA markers.

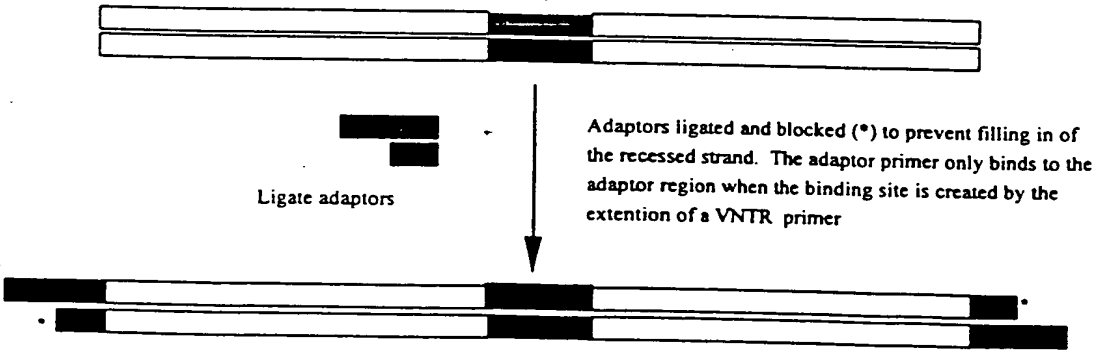
The invention provides a method of selecting genetic elements that are common to one pool of individuals but are absent in a second. An obvious variation on this theme is the selection of genetic elements that are absent in one pool of individuals but are present in a second by judicious selection, during the course of the procedure, of allele duplexes that are either with or without a mis-match.

1. Recreation of the VNTRs with the flanking sequences

Genomic DNA from an individual of the species in which an investigation is to be performed, but not necessarily an individual in that investigation, is digested by one or more restriction endonucleases to fragment the genome.



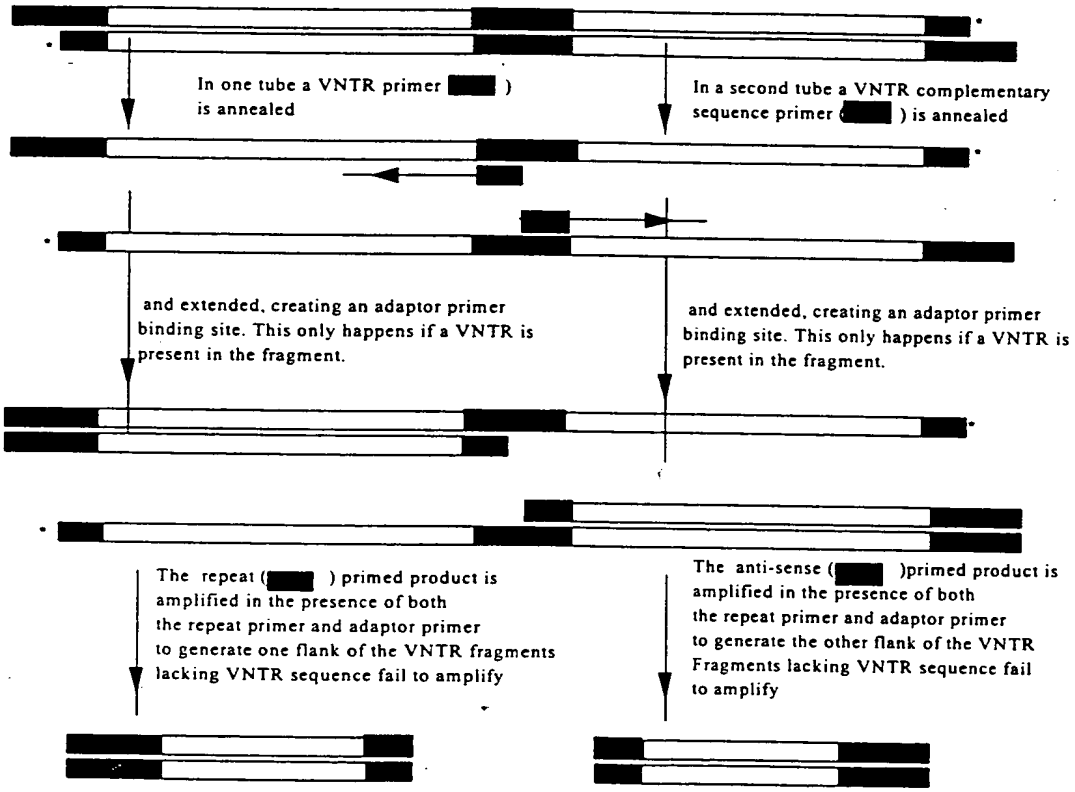
The genomic fragments are ligated to adaptors that possess a recessed 3' end which is blocked chemically, or by other methods, to prevent strand extension in the presence of DNA polymerase. The adaptors may be blocked prior to ligation to the genomic fragments or after the ligation has occurred.



In the absence of strand extension there is no adaptor primer binding site and exponential amplification of the ligated fragment using the adaptor primer is not possible. However, when a chosen primer, with

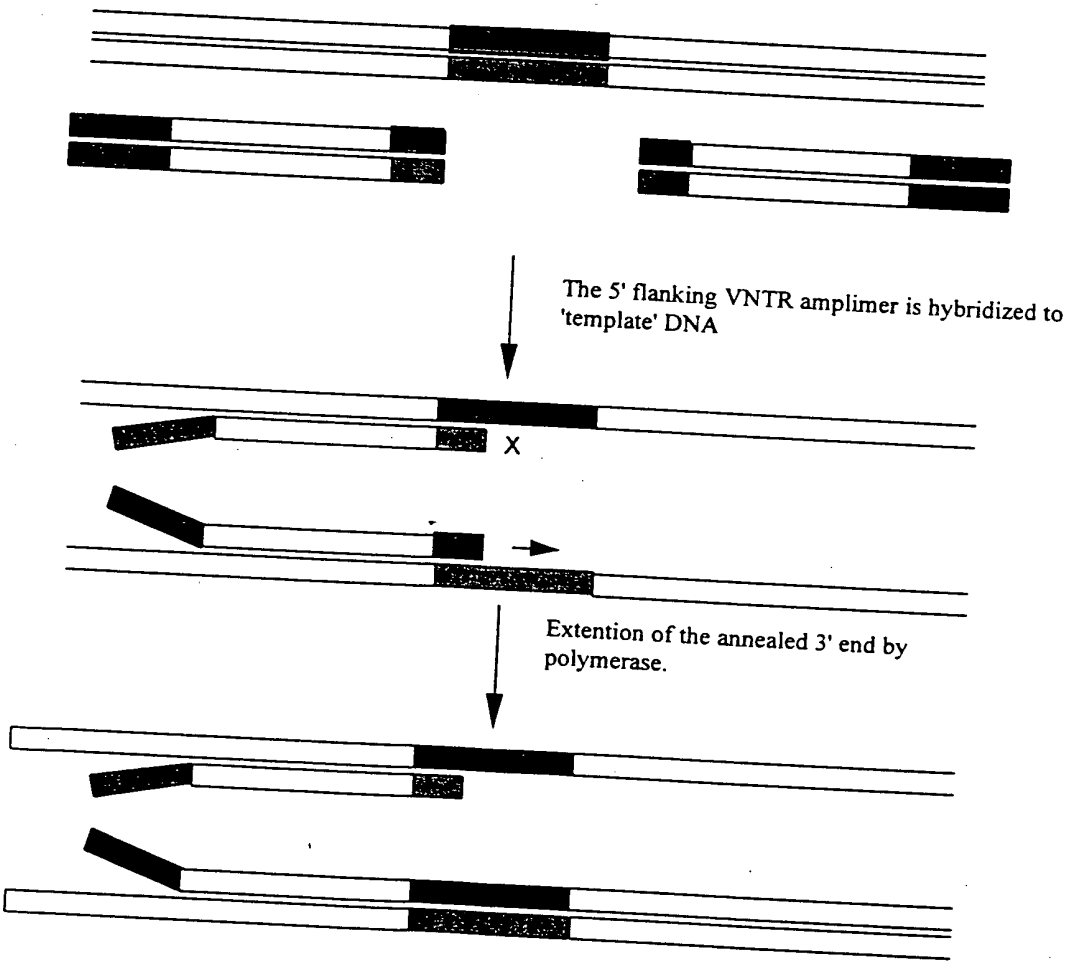
homology to the ligated genomic fragment, anneals and extends to the limit of the adaptor, an adaptor primer binding site is created. All products of amplification, therefore, are both ligated to the chosen adaptor and contain sequence homologous to that of the chosen primer.

In separate reactions the adaptor primer is used with either a VNTR sense primer or a VNTR antisense primer to produce the 3' and 5' flanking sequences of each VNTR, respectively, by PCR, NASBA, or other methods.

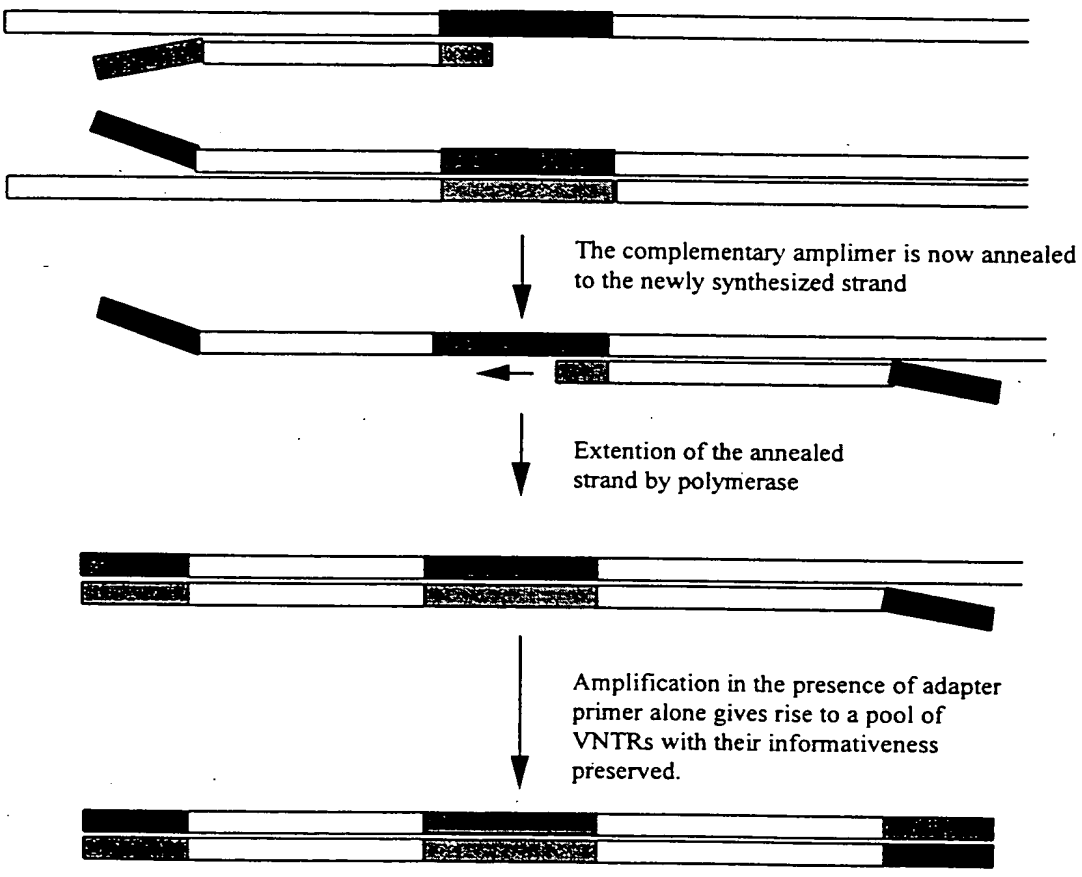


The amplimers are purified from the VNTR primers by physical, chemical or enzymic means and size selection is performed to obtain amplimers of optimal size. The purified amplimers are then used to generate each VNTR allele present in the genomic DNA of one or more individuals manifesting a particular trait of interest. Where a pool of

individuals manifesting a particular trait is to be investigated, the VNTR alleles of each of those individuals may be generated from genomic DNA in isolation using separate reactions, or they may be generated collectively using the pooled genomes of those individuals. This may be achieved by hybridization of genomic DNA of the individuals manifesting a particular hereditary trait to the amplimers representing the 5' region of each VNTR, created using the VNTR antisense primer. The hybridization is incubated with DNA polymerase. Providing an adaptor has been chosen which lacks homology to genomic sequence of any individual under investigation, only one strand of each amplimer is capable of extending in the presence of DNA polymerase.



Addition of the 3' amplimers, generated by the VNTR sense primer, with thermal cycling achieves successive rounds of strand dissociation and reannealing with the generation of each VNTR allele as present in the genomic DNA of those individuals under investigation.



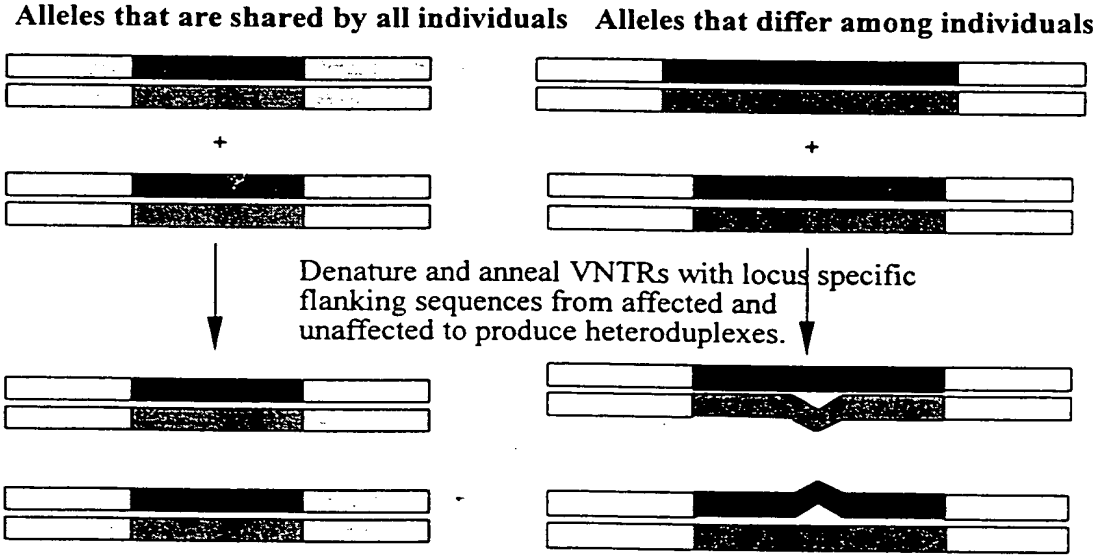
It is also possible to generate the full length VNTRs, with their flanking sequences, from the genomic DNA of individuals manifesting a particular trait by:

(i) hybridization of genomic DNA of individuals manifesting a particular hereditary trait to amplimers representing the 3' region of each VNTR, created using the VNTR sense primer. The hybridization is incubated with DNA polymerase after which the antisense primed amplimers are added;

(ii) hybridization of genomic DNA of individuals manifesting a particular hereditary trait to both amplimer pools, simultaneously, created by the sense and antisense VNTR primers. The hybridization is then incubated with DNA polymerase.

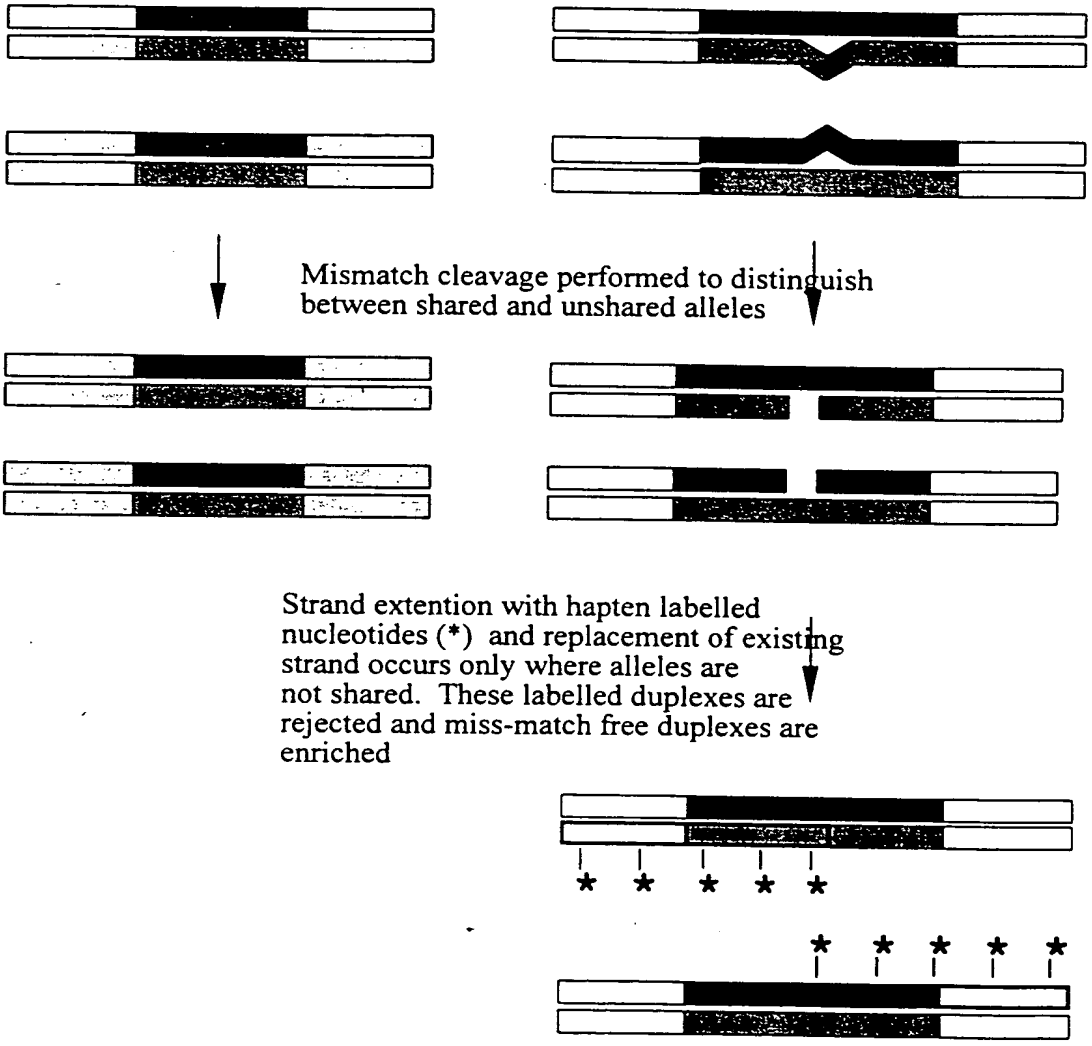
2. Mis-match discrimination of the generated VNTRs with flanking sequences from individuals under investigation.

The VNTR alleles with flanking sequences generated from the genomic DNAs of individuals manifesting a particular trait, synthesized individually or collectively, are hybridized. If different alleles exist at any locus in the genomes of those individuals mis-match containing heteroduplex formation may occur.

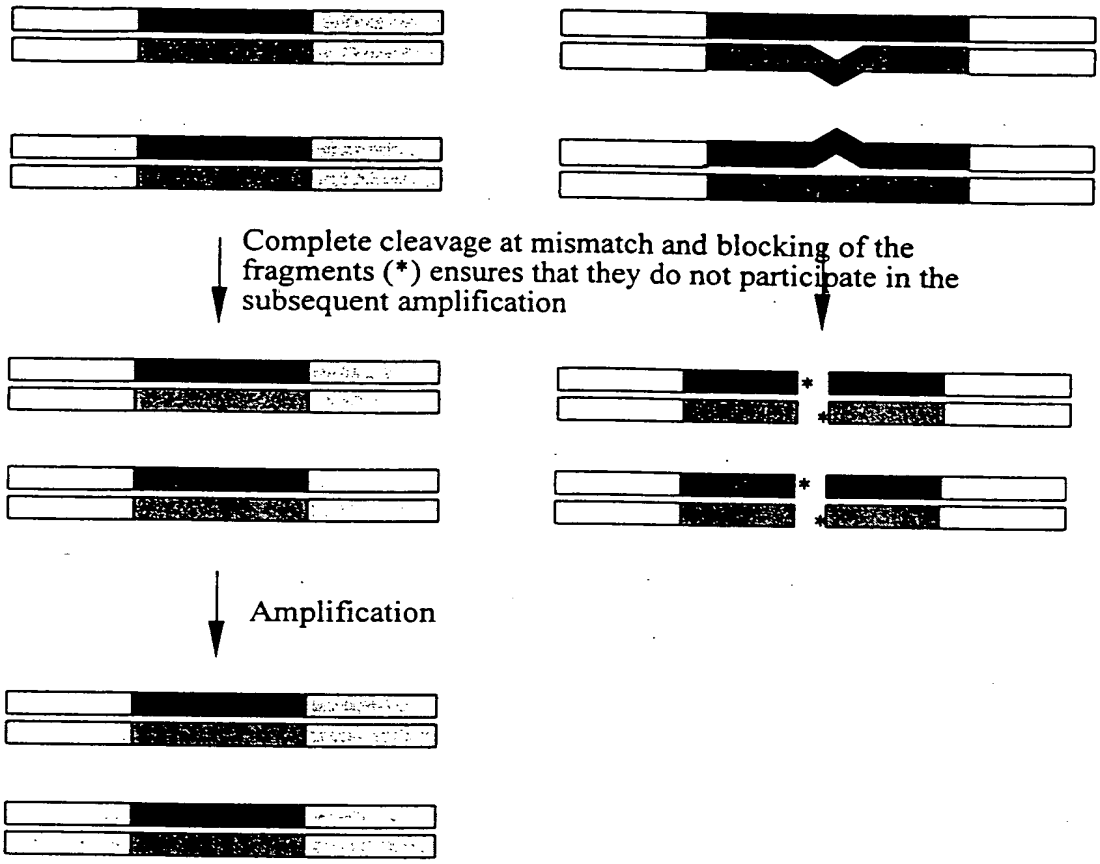


Mis-match containing heteroduplexes are discarded by mis-match discrimination resulting in the enrichment of those alleles common to all individuals manifesting the trait. In addition, this procedure removes the spurious products of amplification resulting from miss-priming events or reaction contamination since these spurious products will not be constant among all individuals. The mis-match discrimination may be achieved by:

(i) partial digestion of the mis-match containing heteroduplexes and incubation with a labelled dNTP. Labelled duplexes that contained mis-matches may be bound to a support from which the mis-match free duplexes can be separated;

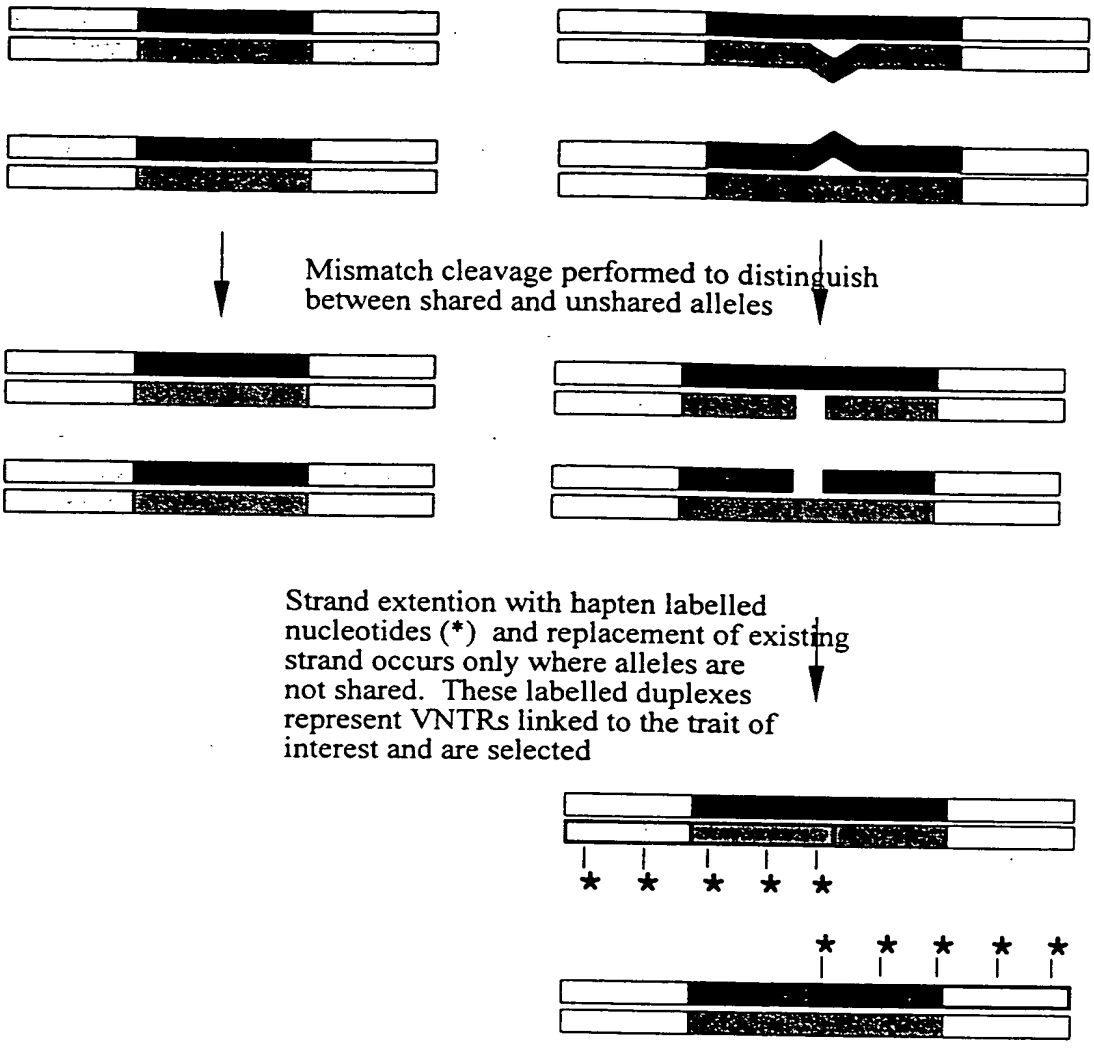


(ii) complete digestion of the mis-match containing heteroduplexes and blockade of the 3' ends to prevent strand extension such that only those strands incorporated in mis-match free duplex formations can achieve exponential amplification and enrichment within the reaction.



The selected duplexes represent VNTR alleles that are common to all individuals manifesting a particular trait of interest. These alleles may be hybridized to the VNTR alleles generated from individuals in which that trait is not present. A portion of the alleles that are common to all individuals manifesting a trait of interest will form mis-match free heteroduplexes with the alleles from those in which the trait is absent. Since these are non-informative for the trait of interest they are rejected. Heteroduplexes containing a mis-match, however, are selected by partial digestion, and incubation with a labelled dNTP and DNA polymerase to allow separation of these alleles from the reaction. The selected heteroduplexes represent VNTRs that are linked to the particular trait of interest. Amplification of these VNTRs from DNA of individuals manifesting the trait yields the informative alleles which constitute a Total

Representation of Alleles Informative for a Trait (TRAIT).



Alternatively, the selected alleles that are common to individuals manifesting a particular trait may be hybridized to an array of fixed alleles of known identity and spatial separation. By mis-match discrimination, the identity of alleles forming mis-match free duplexes can be established.

Th preferred embodiment

(i) Genomic DNA of an individual of the species under

investigation, but not necessarily an individual in that investigation, is fragmented with a single restriction enzyme.

(ii) Adaptors with a recessed 3' end are ligated to all genomic fragments.

(iii) All 3' ends are terminated by incubation of the ligated fragments with ddNTPs in the presence of Taq DNA polymerase.

(iv) The ligated fragments are purified from the ddNTPs and amplified in reactions containing:

a) adaptor primer and a $(CA)_nD$ primer, where D=G+A+T;

b) adaptor primer and an $(AC)_nB$ primer, where B=G+T+C;

c) adaptor primer and a $(GU)_nH$ primer, where H=A+T+C;

d) adaptor primer and a $(UG)_nV$ primer, where V=G+A+C.

The products of amplification result from genomic fragments that successfully ligate to the chosen adaptor and contain a VNTR with homology to the chosen primer.

(v) digestion of the $(GU)_nH$ primed products and the $(UG)_nV$ primed products with uracil-DNA glycosylase followed by Exonuclease 1 to remove all VNTR sequences and excess VNTR primer.

(vi) digestion of $(CA)_nD$ primed products and $(AC)_nB$ primed products with Exonuclease 1 to remove all excess VNTR primer. Size selection may be performed to obtain products of an optimal range of molecular weights.

(vii) Hybridization of equal amounts of $(GU)_nH$ primed products and the $(UG)_nV$ primed products with an equimolar quantity of pooled genomic DNAs derived from individuals manifesting a particular trait of interest. These DNAs act as template for generation of VNTR alleles contained within them.

(viii) Incubation of the hybridized products with Taq DNA polymerase to achieve strand extension of all annealed 3' ends.

(ix) Addition of equal amounts of $(CA)_nD$ primed products and

$(AC)_nB$ primed products to the reaction and generation of VNTR alleles from the 'genomic template' by thermal cycling in the presence of Taq DNA polymerase.

(x) Purification of the generated VNTR alleles followed by strand dissociation and reannealing under stringent conditions.

(xi) Partial digestion with S1 nuclease of mis-match containing heteroduplexes that result from hybridization of VNTR alleles to spurious products of amplification, or hybridization of VNTR alleles that differ among the individuals under investigation manifesting a particular trait of interest.

(xii) Incubation of the hybridized alleles with biotin dUTP in the presence of Taq DNA polymerase such that the digested heteroduplexes incorporate biotin.

(xiii) Separation of biotin containing molecules from the reaction by passage over streptavidin bound to a solid support. The remaining mis-match free duplexes are selected.

(xiv) Repetition of hybridization, digestion and separation of heteroduplex molecules. This enriches the reaction in VNTR alleles that are common to all manifesting the particular trait of interest and removes any spurious products of amplification.

(xv) Hybridization of the selected VNTR alleles, that are common to all individuals manifesting a particular trait, to the VNTR alleles of individuals in which the trait is absent that have been generated from their genomic DNAs in a method bearing similarity, wholly or in part, with (i) to (ix).

(xvi) Partial digestion with S1 nuclease of mis-match containing heteroduplexes followed by incubation with biotin-dUTPs such that the digested heteroduplexes incorporate biotin.

(xvii) Selection of biotin containing molecules from the reaction by passage over streptavidin bound to a solid support. These selected heteroduplexes represent VNTRs that are linked to the particular trait of

interest. The informative allele of these VNTRs, with respect to the trait of interest, can be established from the pool of alleles common to all individuals manifesting that trait by various methods including but not limited amplification using VNTR specific primers.

A second embodiment

(i) Genomic DNA of an individual of the species under investigation, but not necessarily an individual in that investigation, is fragmented with a single restriction enzyme.

(ii) Adaptors with a recessed 3' end are ligated to all genomic fragments.

(iii) All 3' ends are terminated by incubation of the ligated fragments with ddNTPs in the presence of Taq DNA polymerase.

(iv) The ligated fragments are purified from the ddNTPs and amplified in reactions containing:

a) adaptor primer and a $(CA)_nD$ primer, where D=G+A+T;

b) adaptor primer and an $(AC)_nB$ primer, where B=G+T+C;

c) adaptor primer and a $(GU)_nH$ primer, where H=A+T+C;

d) adaptor primer and a $(UG)_nV$ primer, where V=G+A+C.

The products of amplification result from genomic fragments that successfully ligated to the chosen adaptor and contain a VNTR with homology to the chosen primer.

(v) digestion of the $(GU)_nH$ primed products and the $(UG)_nV$ primed products with uracil-DNA glycosylase followed by exonuclease 1 to remove all VNTR sequences and excess VNTR primer.

(vi) digestion of $(CA)_nD$ primed products and $(AC)_nB$ primed products with exonuclease 1 to remove all excess VNTR primer. Size selection may be performed to obtain products of an optimal range of molecular weights.

(vii) Hybridization of equal amounts of $(GU)_nH$ primed products

and the $(UG)_nV$ primed products with an equimolar quantity of pooled genomic DNAs derived from individuals manifesting a particular trait of interest.

(viii) Incubation of the hybridized products with Taq DNA polymerase to achieve strand extension of all annealed 3' ends.

(ix) Addition of equal amounts of $(CA)_nD$ primed products and $(AC)_nB$ primed products to the reaction and generation of VNTR alleles from 'genomic template' by thermal cycling in the presence of Taq DNA polymerase.

(x) Purification of the generated VNTR alleles followed by strand dissociation and reannealing under stringent conditions.

(xi) Partial digestion with S1 nuclease of mis-match containing heteroduplexes that result from hybridization of VNTR alleles to spurious products of amplification, and hybridization of VNTR alleles that differ among the individuals under investigation manifesting a particular trait of interest.

(xii) Incubation of the hybridized alleles with biotin dUTP in the presence of Taq DNA polymerase such that the digested heteroduplexes incorporate biotin.

(xiii) Separation of biotin containing molecules from the reaction by passage over streptavidin bound to a solid support. The remaining mis-match free duplexes are selected.

(xiv) Repetition of hybridization, digestion and separation of heteroduplex molecules. This enriches the reaction in VNTR alleles that are common to all manifesting the particular trait of interest and removes any spurious products of amplification.

(xv) Hybridization of the alleles common to all individuals manifesting a particular trait of interest to a fixed array of all possible VNTR alleles, or subset thereof, on a solid support of known identity and spatial separation under stringent conditions.

(xvi) Partial digestion with S1 nuclease of mis-match containing heteroduplexes followed stringent washing of the array to remove all unbound and digested strands.

(xvii) Detection of alleles forming mis-match free duplexes by observation of fluorescence, providing the VNTR alleles fixed to the array are appropriately labelled. The identity of these alleles is established if the identity and spatial separation of the alleles fixed on the array is known.

A third embodiment

(i) VNTR alleles are generated by means other than processes of amplification of fragmented and ligated genomic DNA with adaptor primer and VNTR primer, hybridization of the generated products to genomic 'template' DNAs of individuals manifesting a particular trait, and generation of the respective VNTR alleles from those template DNAs. These may include but are not limited to:

a) amplification of VNTRs from genomic or synthetic DNA using primers specific to the flanking regions of each VNTR in individual reactions;

b) amplification of VNTRs from genomic or synthetic DNA using a multiplex system, thereby allowing amplification of multiple VNTRs on mass using adapted VNTR specific primers;

c) amplification of VNTRs from genomic or synthetic DNA using an endonuclease that cleaves in or about VNTR sequences such that adaptors may be ligated to the digested DNA and used for amplification of the VNTR alleles;

d) generation of a pool of VNTRs from individuals manifesting a particular trait by processes of subtraction with those in which the trait is absent.

(ii) Purification of the generated VNTR alleles followed by strand dissociation and reannealing under stringent conditions.

(iii) Partial digestion with S1 nuclease of mis-match containing heteroduplexes that result from hybridization of VNTR alleles to spurious products of amplification, and hybridization of VNTR alleles that differ among the individuals under investigation manifesting a particular trait of interest.

(iv) Incubation of the hybridized alleles with biotin dUTP in the presence of Taq DNA polymerase such that the digested heteroduplexes incorporate biotin.

(v) Separation of biotin containing molecules from the reaction by passage over streptavidin bound to a solid support. The remaining mis-match free duplexes are selected.

(vi) Repetition of hybridization, digestion and separation of heteroduplex molecules. This enriches the reaction in VNTR alleles that are common to all manifesting the particular trait of interest and removes any spurious products of amplification.

(xv) Hybridization of the alleles common to all individuals manifesting a particular trait of interest to a fixed array of all possible VNTR alleles, or subset thereof, on a solid support of known spatial separation under stringent conditions.

(xvi) Partial digestion with S1 nuclease of heteroduplexes followed stringent washing of the array to remove all unbound strands and digested molecules.

(xvii) Detection of alleles forming mis-match free duplexes by observation of fluorescence, providing the VNTR alleles fixed to the array are appropriately labelled. The identity of these alleles is established if the identity and spatial separation of the alleles fixed on the array is known.

Those that are skilled in the art will appreciate that there are several methods of differentiating mis-match containing duplexes from those that are free of mis-matches, either in solution or on an array. The methods described in the above embodiments represent only one of these

methods.

Those that are skilled in the art will appreciate that the invention is equally well suited any type of VNTR including but not restricted to dinucleotide repeats e.g.$(CA)_n$ and $(GT)_n$, trinucleotide repeats e.g.$(AAT)_n$, $(AGC)_n$, $(AGG)_n$, $(CAC)_n$, $(CCG)_n$ and $(CTT)_n$, and tetranucleotide repeats e.g.$(CCTA)_n$, $(CTGT)_n$, $(CTTT)_n$.$(TAGG)_n$, $(TCTA)_n$, and $(TTCC)_n$. In addition, the invention may be applied to simple organism microsatellites that include, but are not limited to, (AT), (CC), (CT) and (GA) rich tracts of repetitive motifs.

Those that are skilled in the art will appreciate that polymorphic alleles, other than those of VNTRs, may be used with the invention to produce alleles that are free of spurious products of amplification and are common to all individuals manifesting a particular trait. These polymorphic alleles may be hybridized to a fixed array of all possible alleles, or subset thereof, or to a pool of alleles derived from individuals in which that trait is absent. By mis-match discrimination using S1 nuclease and incorporation of biotin, those alleles linked and informative for a trait can be identified.

Those that are skilled in the art will appreciate that alleles from the genome of a single individual, or more than one individual, of unknown phenotype and genotype may be amplified with fidelity, removing the spurious products of amplification by mis-match discrimination, and hybridized to a fixed array of alleles, or to a pool of alleles in solution, in order assign a genotype or a phenotype to that individual.

Those that are skilled in the art will appreciate that mis-match discrimination may be performed using enzymes or chemicals other than S1 nuclease. These alternatives include but are not limited to Mung Bean nuclease, mutation detection proteins (e.g. Mut S), recombinases (e.g. Endonuclease VII), osmium tetroxide and hydroxylamine.

Those that are skilled in the art will appreciate that haptens other than biotin, using chemicals with affinity to the chosen hapten bound

on a solid support, may be used to achieve separation of duplexes with or without mis-matches after mis-match discrimination.

The invention differs fundamentally from all previous techniques since genomic fragments are generated that do not reflect the polymorphic variation at the locus from which they were derived. Furthermore, these fragments need not be generated from an individual in a particular investigation, but may be from any individual of the appropriate species. However, hybridization of these fragments to genomic 'template' DNA of an individual subject to investigation and mis-match discrimination permits amplification, with fidelity, of alleles within that genomic template whilst overcoming the problems of generation of spurious products that are a feature of other PCR-based methods. Since the invention preserves each VNTR allele with its flanking sequences, these alleles remain highly informative. In this respect the invention is unique. Furthermore, this novel method of generating VNTRs is rapid, inexpensive, has no requirement for prior knowledge of sequence, and has no requirement for elaborate equipment, it is of immense importance obviating the high investment of time and money that is currently required for isolation of VNTRs. Consequently, the application of technologies dependant on the availability of VNTR in species in which none have been isolated will be possible where previously this was unfeasible. The ability to generate large numbers of VNTRs from all species quickly, efficiently, cheaply and with fidelity is a considerable contribution of the present invention to workers in the to the biomedical field.

In summary, the invention involves a novel method of generating VNTRs encompassing restriction endonuclease digestion of DNA, ligation of the fragments to adaptors and, by introduction of a primer with sequence homology to a chosen VNTR, amplifying only those fragments that are flanked by a chosen endonuclease restriction enzyme site and a VNTR. These fragments are not representative of the alleles of

each VNTR and need not be generated from any specific individual under investigation. Hybridization of these fragments with genomic DNA of the individuals under investigation recreates the intact VNTR alleles with flanking sequence, as they occur in the genome. This in itself constitutes a major step in the ability of workers in the biomedical fields to generate quickly, efficiently, cheaply and with fidelity VNTRs in all species for purposes reliant on the availability of VNTRs, including but not confined to DNA fingerprinting and linkage analysis. The incorporation of a mis-match discrimination procedure overcomes the problems of miss-priming and generation of spurious products by reaction contamination and subtle variation in reaction conditions, that are to the detriment of all PCR-based technologies, and allows exclusion of alleles that are not common to all individuals under investigation that manifest a particular trait. A second round of mis-match discrimination removes un-informative alleles that are present in the genomes of individuals that do not manifest the trait. This procedure is designated a Total Representation of Alleles that are Informative for a Trait (TRAIT). The invention, therefore, has significant advantages over previous methods, embracing the speed of analysis of AFLP, RDA and RAPD, and the high polymorphism detection rate of linkage analysis, but negating the need for DNA from closely related individuals and for paternity testing. The invention also overcomes fundamental problems that are a feature of PCR based technologies, including miss-priming and generation of spurious products through reaction contamination and subtle variations in the conditions of reaction. Furthermore, there is no requirement for expensive equipment or elaborate statistical computer software. The analysis will give rise to alleles that are both linked and informative, being present in individuals manifesting the trait of interest but absent in those individuals that lack the trait. In this respect, the invention is unchallenged in its superiority over all other methods.

The invention allows concomitant detection of polymorphisms at multiple loci by simultaneous comparison of simple or complex genomes from multiple individuals and differs fundamentally from all other techniques that have been previously employed. The invention represents a major advance in the ability of workers in the biomedical fields to generate VNTRs from the genomes of any species quickly, efficiently, cheaply and with fidelity in addition to screening complex genomes for polymorphisms co-segregating with hereditary traits. Application of this procedure will therefore facilitate the development of markers for genetic screening for hereditary disease, or advantageous monogenic or polygenic traits in all organisms.

Examples of how the invention may be applied

The following illustrations represent examples of how the invention may be applied without inferring any limitation to scope of the invention or any limitation to the different ways in which the invention may be applied.

Example 1

In this example the $(CA)_n/(GT)_n$ VNTRs are generated from genomic DNA fragmented by a single restriction enzyme. The amplified products are hybridized to two pools of genomic DNA. One pool is derived from individuals manifesting a particular hereditary trait. The second pool is derived from individuals in which that trait is absent. Mis-match discrimination is performed solely by partial digestion of mis-match containing heteroduplexes.

(i) restriction digestion of genomic DNA using Rsa1 buffered appropriately at 37°C for several hours to yield genomic fragments.

(ii) Annealing of the sense and antisense oligonucleotides that comprise the adaptor (a 48mer and 12mer respectively). This may be

achieved by mixing equimolar quantities of each oligo in ligase buffer and heating to 50°C followed by gradual cooling to 10°C over 1 hour.

(iii) Ligating the adaptor to the genomic fragments by addition of the fragments to the buffered adaptor and incubating at 15°C over night in the presence of T4 DNA ligase.

(iv) Removal of the ligase buffer and incubation of the ligated fragments with dideoxynucleotide triphosphates (ddNTPs) in the presence of Taq DNA polymerase appropriately buffered at 72°C for 1 hour. Complete termination of the recessed 3' ends of the adaptor can be assessed by PCR of an aliquot of the reaction with adaptor primer which should fail to yield any products of amplification.

(v) Purification of the ligated fragments from the ddNTPs and amplification of those ligated fragments containing a $(CA)_n/(GT)_n$ VNTR under appropriate conditions with either a $(CA)_n$ primer and the adaptor primer, or a $(GT)_n$ primer and the adaptor primer. Over amplification of products should be avoided to limit the opportunity for formation of concatamers. Using PCR a total of about 25 cycles may be appropriate, though NASBA or other methods may be employed to achieve the same goal.

(vi) The $(CA)_n$ primers and the $(GT)_n$ primers are destroyed by physical separation of molecules based on molecular weight, electrophoretic mobility, or by digestion of single stranded DNA using an appropriate enzymes such as Exonclease 1 appropriately buffered at 37°C.

(vii) hybridization of the $(CA)_n$ primed products to an equimolar quantity of the pooled genomic DNAs of all individuals under investigation manifesting the particular trait of interest. Hybridization should proceed under stringent conditions of high temperature and low salt concentration. Strand dissociation at 98°C for 5 minutes followed by a step-wise decrease in temperature to 78°C over several hours in solution of limited volume containing 0.2M NaCl may be appropriate.

(viii) incubation of the hybridization components with Taq DNA polymerase appropriately buffered at 72°C to achieve extension of annealed strands.

(ix) addition of an amount of $(GT)_n$ primed products, similar to that of the $(CA)_n$ primed products, and adaptor primer to the reaction followed by amplification of products to yield VNTR alleles faithfully reproduced from the genomic templates of individuals manifesting the trait of interest. This may be performed by PCR, NASBA or other methods.

(x) dissociation of the amplified VNTR alleles by incubation at 98°C for 5 minutes then step-wise reduction in temperature to 78°C over several hours in a limited reaction volume containing 0.2M NaCl. This allows the alleles at each locus to hybridize to the alleles derived from other individuals resulting in mis-match free duplex formation where alleles are in common, or duplexes with a mis-match where alleles differ among individuals.

(xi) Incubation of the hybridized strands with S1 nuclease under appropriate conditions to achieve partial digestion of all mis-match containing duplex molecules.

(xii) incubation of the reaction components with biotin-dUTP containing dNTPs in the presence of Taq DNA polymerase appropriately buffered at 72°C for 1 hour to achieve extension of the 3' ends of all digested strands such that mis-match containing duplexes incorporate biotin.

(xiii) physical separation of biotin containing duplexes due by binding to streptavidin on a rigid support.

(xiv) repeated dissociation and reannealing of the reaction components followed by heteroduplex digestion and biotin incorporation to permits enrichment of alleles common to all individuals manifesting the trait of interest.

(xv) amplification of the VNTR alleles of the pool of individuals in

which the trait of interest in absent, as described in (i) to (ix).

(xvi) hybridization of the enriched VNTR alleles forming mis-match free duplexes, that are common to all individuals manifesting a the trait of interest, with the VNTR alleles derived from those individuals in which the trait is absent. This may be achieved by strand dissociation at 98°C for 5 minutes then step-wise reduction in temperature over several hours to 78°C in a limited reaction volume containing 0.2M NaCl.

(xvii) partial digestion of the mis-match containing duplexes with S1 nuclease under appropriate conditions.

(xviii) incubation of the reaction components with biotin-dUTP containing dNTPs in the presence of Taq DNA polymerase appropriately buffered at 72°C for 1 hour.

(xix) separation of the biotin containing duplex molecules formed by the association of alleles common to individuals manifesting the trait of interest that are absent in those in which the trait is not present. The biotin containing molecules are selected and represent VNTRs that are linked to the trait of interest. The informative alleles of these VNTRs can be amplified from the DNAs of individuals manifesting the trait of interest.

Example 2

In this example the $(CA)_n/(GT)_n$ VNTRs are generated from genomic DNA fragmented by a single restriction enzyme. The amplified products are hybridized to two pools of genomic DNA. One pool is derived from individuals manifesting a particular hereditary trait. The second pool is derived from individuals in which that trait is absent. Mis-match discrimination among all individuals manifesting the particular trait of interest is performed by complete digestion of the mis-match containing duplexes followed by enrichment of mis-match free duplexes by amplification. The amplified products represent alleles that a common to all individuals manifesting the trait of interest. This is a less preferred

embodiment since the amplification of the selected alleles may yield spurious products by miss-priming events or reaction contamination. However, hybridization of these alleles to the alleles derived from individuals in which the trait is absent followed by partial digestion of any mis-match containing duplexes may be performed.

(i) restriction digestion of genomic DNA using Rsa1 buffered appropriately at 37°C for several hours to yield genomic fragments.

(ii) Annealing of the sense and antisense oligonucleotides that comprise the adaptor (a 48mer and 12mer respectively). This may be achieved by mixing equimolar quantities of each oligo in ligase buffer and heating to 50°C followed by gradual cooling to 10°C over 1 hour.

(iii) Ligating the adaptor to the genomic fragments by addition of the fragments to the buffered adaptor and incubating at 15°C over night in the presence of T4 DNA ligase.

(iv) Removal of the ligase buffer and incubation of the ligated fragments with dideoxynucleotide triphosphates (ddNTPs) in the presence of Taq DNA polymerase appropriately buffered at 72°C for 1 hour. Complete termination of the recessed 3' ends of the adaptor can be assessed by PCR of an aliquot of the reaction with adaptor primer which should fail to yield any products of amplification.

(v) Purification of the ligated fragments from the ddNTPs and amplification of those ligated fragments containing a $(CA)_n/(GT)_n$ VNTR under appropriate conditions with either a $(CA)_n$ primer and the adaptor primer, or a $(GT)_n$ primer and the adaptor primer. Over amplification of products should be avoided to limit the opportunity for formation of concatamers. Using PCR a total of about 25 cycles may be appropriate, though NASBA or other methods may be employed to achieve the same goal.

(vi) The $(CA)_n$ primers and the $(GT)_n$ primers are destroyed by physical separation of molecules based on molecular weight,

electrophoretic mobility, or by digestion of single stranded DNA using an appropriate enzymes such as Exonclease 1 appropriately buffered at 37°C.

(vii) hybridization of the $(CA)_n$ primed products to an equimolar quantity of the pooled genomic DNAs of all individuals under investigation manifesting the particular trait of interest. Hybridization should proceed under stringent conditions of high temperature and low salt concentration. Strand dissociation at 98°C for 5 minutes followed by a step-wise decrease in temperature to 78°C over several hours in solution of limited volume containing 0.2M NaCl may be appropriate.

(viii) incubation of the hybridization components with Taq DNA polymerase appropriately buffered at 72°C to achieve extension of annealed strands.

(ix) addition of an amount of $(GT)_n$ primed products, similar to that of the $(CA)_n$ primed products, and adaptor primer to the reaction followed by amplification of products to yield VNTR alleles faithfully reproduced from the genomic templates of individuals manifesting the trait of interest. This may be performed by PCR, NASBA or other methods.

(x) dissociation of the amplified VNTR alleles by incubation at 98°C for 5 minutes then step-wise reduction in temperature to 78°C over several hours in a limited reaction volume containing 0.2M NaCl. This allows the alleles at each locus to hybridize to the alleles derived from other individuals resulting in mis-match free duplex formation where alleles are in common, or duplexes containing a mis-match where alleles differ among individuals.

(xi) Incubation of the hybridized strands with S1 nuclease under appropriate conditions to achieve complete digestion of all mis-match duplexes.

(x) purification of the reaction components and incubation with ddUTP containing dNTPs in the presence of Taq DNA polymerase at 72°C for 1 hour to block all 3' ends of the annealed strands.

(xi) purification of the reaction components and amplification of mis-match free duplexes by PCR, NASBA, or other methods.

(xii) repeated dissociation and reannealing of the reaction components followed by heteroduplex digestion, and incubation with ddUTP containing dNTPs to achieve enrichment of alleles common to all individuals manifesting the trait of interest by PCR, NASBA or other methods.

(xiii) amplification of the VNTR alleles of the pool of individuals in which the trait of interest in absent, as described in (i) to (ix).

(xiv) hybridization of the enriched VNTR alleles that are common to all individuals manifesting a the trait of interest with the VNTR alleles derived from those individuals in which the trait is absent. This may be achieved by strand dissociation at 98°C for 5 minutes then step-wise reduction in temperature over several hours to 78°C in a limited reaction volume containing 0.2M NaCl.

(xv) partial digestion of the mis-match containing duplexes with S1 nuclease under appropriate conditions.

(xviii) incubation of the reaction components with biotin-dUTP containing dNTPs in the presence of Taq DNA polymerase appropriately buffered at 72°C for 1 hour.

(xix) separation of the biotin containing duplex molecules formed by the association of alleles that differ between individuals manifesting the trait of interest and those in which the trait is absent. The biotin containing molecules are selected and represent VNTRs that are linked to the trait of interest. The informative alleles of these VNTRs can be amplified from the DNAs of individuals manifesting the trait of interest.

Example 3

In this example the $(CA)_n/(GT)_n$ VNTRs are generated from genomic DNA fragmented by a single restriction enzyme. The amplified

products are hybridized to the genomic 'template' DNA of a single individual manifesting a particular trait of interest to yield the VNTR alleles represented in that genomic 'template' DNA. By hybridization of the amplified VNTR alleles to a sample of the same genomic 'template' DNA, partial digestion of any mis-match containing duplexes, and incubation of the products with biotin-dUTP, the spurious products of amplification due to reaction contamination and miss-priming events are removed. The non-biotin containing alleles forming mis-match free duplexes with the genomic 'template' DNA are hybridized to the pooled VNTR alleles of individuals that do not manifest the trait of interest. Heteroduplexes containing mis-matches formed in this hybridization are selected by partial digestion and incubation of biotin-dUTP in the presence of Taq DNA polymerase.

(i) restriction digestion of genomic DNA using Rsa1 buffered appropriately at 37°C for several hours to yield genomic fragments.

(ii) Annealing of the sense and antisense oligonucleotides that comprise the adaptor (a 48mer and 12mer respectively). This may be achieved by mixing equimolar quantities of each oligo in ligase buffer and heating to 50°C followed by gradual cooling to 10°C over 1 hour.

(iii) Ligating the adaptor to the genomic fragments by addition of the fragments to the buffered adaptor and incubating at 15°C over night in the presence of T4 DNA ligase.

(iv) Removal of the ligase buffer and incubation of the ligated fragments with dideoxynucleotide triphosphates (ddNTPs) in the presence of taq DNA polymerase appropriately buffered at 72°C for 1 hour. Complete termination of the recessed 3' ends of the adaptor can be assessed by PCR of an aliquot of the reaction with adaptor primer which should fail to yield any products of amplification.

(v) Purification of the ligated fragments from the ddNTPs and amplification of those ligated fragments containing a $(CA)_n/(GT)_n$ VNTR under appropriate conditions with either a $(CA)_n$ primer and the adaptor

primer, or a $(GT)_n$ primer and the adaptor primer. Over amplification of products should be avoided to limit the opportunity for formation of concatamers. Using PCR a total of about 25 cycles may be appropriate, though NASBA or other methods may be employed to achieve the same goal.

(vi) The $(CA)_n$ primers and the $(GT)_n$ primers are destroyed by physical separation of molecules based on molecular weight, electrophoretic mobility, or by digestion of single stranded DNA using an appropriate enzymes such as Exonclease 1 appropriately buffered at 37°C.

(vii) hybridization of the $(CA)_n$ primed products to an equimolar quantity of the genomic 'template' DNA of the single individual under investigation manifesting the particular trait of interest. Hybridization should proceed under stringent conditions of high temperature and low salt concentration. Strand dissociation at 98°C for 5 minutes followed by a step-wise decrease in temperature to 78°C over several hours in solution of limited volume containing 0.2M NaCl may be appropriate.

(viii) incubation of the hybridization components with taq DNA polymerase appropriately buffered at 72°C to achieve extension of annealed strands.

(ix) addition of an amount of $(GT)_n$ primed products, similar to that of the $(CA)_n$ primed products, and adaptor primer to the reaction followed by amplification of products to yield VNTR alleles faithfully reproduced from the genomic template of the individual manifesting the trait of interest. This may be performed by PCR, NASBA or other methods.

(x) addition of the VNTR alleles to a molar excess of the same genomic 'template' DNA from which the alleles were derived followed by dissociation of all strands by incubation at 98°C for 5 minutes then step-wise reduction in temperature to 78°C over several hours in a limited reaction volume containing 0.2M NaCl. This allows the alleles to hybridize to the appropriate locus on the genomic DNA. Any mis-match containing

duplexes that may occur result from miss-priming events or reaction contamination.

(xi) partial digestion of the mis-match containing duplexes using S1 nuclease under appropriate conditions.

(xii) incubation of the reaction components with biotin-dUTP in the presence of Taq DNA polymerase appropriately buffered at 72°C for 1 hour.

(xiii) separation of the biotin-containing molecules by binding to streptavidin on a solid support.

(xiv) repeated hybridization of the VNTR alleles to the same genomic 'template' DNA followed by partial digestion of mis-match containing duplex molecules and incubation with biotin-dUTP for removal of all spurious products of amplification by binding to streptavidin on a solid support.

(xv) amplification of the VNTR alleles derived from the pool of individuals in which the trait of interest in absent, as described in (i) to (ix).

(xvi) hybridization of the VNTR alleles, that represent products which have amplified with fidelity from the individual manifesting the particular trait of interest, to the pool of VNTR alleles derived from individuals in which the trait is not present. This may be achieved by strand dissociation at 98°C for 5 minutes then step-wise reduction in temperature over several hours to 78°C in a limited reaction volume containing 0.2M NaCl.

(xvii) partial digestion of the mis-match containing duplexes with S1 nuclease under appropriate conditions.

(xix) incubation of the reaction components with biotin-dUTP containing dNTPs in the presence of Taq DNA polymerase appropriately buffered at 72°C for 1 hour.

(xx) separation of the biotin containing duplex molecules formed by the association of alleles that differ between individual manifesting the

trait of interest and those in which the trait is absent. The biotin containing molecules are selected and may be further amplified. These selected alleles are present in the genome of the individual manifesting the trait of interest but are absent in those in which the trait is not present.

Example 4

In this example the $(CA)_n/(GT)_n$ VNTRs are generated from genomic DNA fragmented by a single restriction enzyme. The amplified products are hybridized to a pool of genomic DNA derived from individuals manifesting a particular hereditary trait. Mis-match discrimination is performed by partial strand dissociation and reannealing, followed by partial digestion of mis-match containing duplexes by S1 nuclease. Digested duplexes are removed by biotin incorporation and binding of streptavidin. The alleles forming mis-match free duplexes are selected and hybridized to an array of VNTR alleles of known identity and spatial separation that are derived from the genomes of individuals in which the trait is absent. Providing the VNTR alleles fixed to the array are appropriately labelled, the identity of the alleles forming mis-match free duplexes with the array may be established.

(i) restriction digestion of genomic DNA using Rsa1 buffered appropriately at 37°C for several hours to yield genomic fragments.

(ii) Annealing of the sense and antisense oligonucleotides that comprise the adaptor (a 48mer and 12mer respectively). This may be achieved by mixing equimolar quantities of each oligo in ligase buffer and heating to 50°C followed by gradual cooling to 10°C over 1 hour.

(iii) Ligating the adaptor to the genomic fragments by addition of the fragments to the buffered adaptor and incubating at 15°C over night in the presence of T4 DNA ligase.

(iv) Removal of the ligase buffer and incubation of the ligated fragments with dideoxynucleotide triphosphates (ddNTPs) in the presence

of taq DNA polymerase appropriately buffered at 72°C for 1 hour. Complete termination of the recessed 3' ends of the adaptor can be assessed by PCR of an aliquot of the reaction with adaptor primer which should fail to yield any products of amplification.

(v) Purification of the ligated fragments from the ddNTPs and amplification of those ligated fragments containing a $(CA)_n/(GT)_n$ VNTR under appropriate conditions with either a $(CA)_n$ primer and the adaptor primer, or a $(GT)_n$ primer and the adaptor primer. Over amplification of products should be avoided to limit the opportunity for formation of concatamers. Using PCR a total of about 25 cycles may be appropriate, though NASBA or other methods may be employed to achieve the same goal.

(vi) The $(CA)_n$ primers and the $(GT)_n$ primers are destroyed by physical separation of molecules based on molecular weight, electrophoretic mobility, or by digestion of single stranded DNA using an appropriate enzymes such as Exonclease 1 appropriately buffered at 37°C.

(vii) hybridization of the $(CA)_n$ primed products to an equimolar quantity of the pooled genomic DNAs of all individuals under investigation manifesting the particular trait of interest. Hybridization should proceed under stringent conditions of high temperature and low salt concentration. Strand dissociation at 98°C for 5 minutes followed by a step-wise decrease in temperature to 78°C over several hours in solution of limited volume containing 0.2M NaCl may be appropriate.

(viii) incubation of the hybridization components with Taq DNA polymerase appropriately buffered at 72°C to achieve extension of annealed strands.

(ix) addition of an amount of $(GT)_n$ primed products, similar to that of the $(CA)_n$ primed products, and adaptor primer to the reaction followed by amplification of products to yield VNTR alleles faithfully reproduced from the genomic templates of individuals manifesting the trait

of interest. This may be performed by PCR, NASBA or other methods.

(x) dissociation of the amplified VNTR alleles by incubation at 98°C for 5 minutes then step-wise reduction in temperature to 78°C over several hours in a limited reaction volume containing 0.2M NaCl. This allows the alleles at each locus to hybridize to the alleles derived from other individuals resulting in mis-match free duplex formation where alleles are in common, or duplexes containing mis-matches where alleles differ among individuals.

(xi) Incubation of the hybridized strands with S1 nuclease under appropriate conditions to achieve partial digestion of all mis-match containing duplex molecules.

(xii) incubation of the reaction components with biotin-dUTP containing dNTPs in the presence of Taq DNA polymerase appropriately buffered at 72°C for 1 hour to achieve extension of the 3' ends of all digested strands such that digested duplexes incorporate biotin.

(xiii) physical separation of biotin containing duplexes due by binding to streptavidin on a rigid support.

(xiv) repeated dissociation and reannealing of the reaction components followed by digestion and biotin incorporation to permits enrichment of alleles common to all individuals manifesting the trait of interest.

(xv) hybridization of the selected alleles forming mis-match free duplexes with the alleles of individuals in which the trait is absent that are fixed to a solid support with known spatial separation. This may be performed by addition of the VNTR alleles from individuals manifesting the trait of interest, that have been selected by mis-match discrimination, to the array of fixed alleles and incubating at 98°C for 5 minutes followed by step-wise decrease in temperature to 78°C over several hours in a limited volume containing 0.2M NaCl held on the array under a coverslip.

(xvi) addition of S1 nuclease and partial digestion of all mis-match

containing heteroduplexes, under appropriate conditions.

(xvii) stringent washing of the array to remove strands that failed to hybridize to the array and duplexes that have been digested This may be achieved by washing with SSC of low molar concentration at high temperature.

(xviii) detection of those alleles that formed mis-match free duplexes with the array and remain intact. This may be achieved by observation of fluorescence providing VNTR alleles bound to the array are appropriately labelled.

Example 5

In this example the $(CA)_n/(GT)_n$ VNTRs are generated from genomic DNA fragmented by a single restriction enzyme. The amplified products are hybridized to a pool of genomic DNA derived from individuals manifesting a particular hereditary trait. Mis-match discrimination is performed by partial strand dissociation and reannealing, followed by partial digestion of mis-match containing duplexes by S1 nuclease. Digested duplexes are removed by biotin incorporation and binding of streptavidin. The alleles forming mis-match free duplexes are selected and hybridized to an array of VNTR alleles of known identity and spatial separation in which all possible alleles are represented, or subset thereof. Providing the VNTR alleles that are fixed to the array are appropriately labelled the identity of the alleles forming mis-match free duplexes with the array may be established.

(i) restriction digestion of genomic DNA using Rsa1 buffered appropriately at 37°C for several hours to yield genomic fragments.

(ii) Annealing of the sense and antisense oligonucleotides that comprise the adaptor (a 48mer and 12mer respectively). This may be achieved by mixing equimolar quantities of each oligo in ligase buffer and heating to 50°C followed by gradual cooling to 10°C over 1 hour.

(iii) Ligating the adaptor to the genomic fragments by addition of the fragments to the buffered adaptor and incubating at 15°C over night in the presence of T4 DNA ligase.

(iv) Removal of the ligase buffer and incubation of the ligated fragments with dideoxynucleotide triphosphates (ddNTPs) in the presence of taq DNA polymerase appropriately buffered at 72°C for 1 hour. Complete termination of the recessed 3' ends of the adaptor can be assessed by PCR of an aliquot of the reaction with adaptor primer which should fail to yield any products of amplification.

(v) Purification of the ligated fragments from the ddNTPs and amplification of those ligated fragments containing a $(CA)_n/(GT)_n$ VNTR under appropriate conditions with either a $(CA)_n$ primer and the adaptor primer, or a $(GT)_n$ primer and the adaptor primer. Over amplification of products should be avoided to limit the opportunity for formation of concatamers. Using PCR a total of about 25 cycles may be appropriate, though NASBA or other methods may be employed to achieve the same goal.

(vi) The $(CA)_n$ primers and the $(GT)_n$ primers are destroyed by physical separation of molecules based on molecular weight, electrophoretic mobility, or by digestion of single stranded DNA using an appropriate enzymes such as Exonclease 1 appropriately buffered at 37°C.

(vii) hybridization of the $(CA)_n$ primed products to an equimolar quantity of the pooled genomic DNAs of all individuals under investigation manifesting the particular trait of interest. Hybridization should proceed under stringent conditions of high temperature and low salt concentration. Strand dissociation at 98°C for 5 minutes followed by a step-wise decrease in temperature to 78°C over several hours in solution of limited volume containing 0.2M NaCl may be appropriate.

(viii) incubation of the hybridization components with Taq DNA polymerase appropriately buffered at 72°C to achieve extension of

annealed strands.

(ix) addition of an amount of $(GT)_n$ primed products, similar to that of the $(CA)_n$ primed products, and adaptor primer to the reaction followed by amplification of products to yield VNTR alleles faithfully reproduced from the genomic templates of individuals manifesting the trait of interest. This may be performed by PCR, NASBA or other methods.

(x) dissociation of the amplified VNTR alleles by incubation at 98°C for 5 minutes then step-wise reduction in temperature to 78°C over several hours in a limited reaction volume containing 0.2M NaCl. This allows the alleles at each locus to hybridize to the alleles derived from other individuals resulting in mis-match free duplexes where alleles are in common, or duplexes containing a mis-match where alleles differ among individuals.

(xi) Incubation of the hybridized strands with S1 nuclease under appropriate conditions to achieve partial digestion of all mis-match containing duplex molecules.

(xii) incubation of the reaction components with biotin-dUTP containing dNTPs in the presence of Taq DNA polymerase appropriately buffered at 72°C for 1 hour to achieve extension of the 3' ends of all digested strands such that digested duplexes incorporate biotin.

(xiii) physical separation of biotin containing duplexes due by binding to streptavidin on a rigid support.

(xiv) repeated dissociation and reannealing of the reaction components followed by heteroduplex digestion and biotin incorporation to permits enrichment of alleles common to all individuals manifesting the trait of interest.

(xv) hybridization of the selected alleles forming mis-match free duplexes with the array of all possible alleles, or subset thereof, fixed to a solid support with known identity and spatial separation. This may be performed by addition of the VNTR alleles from individuals manifesting the

trait of interest, that have been selected by mis-match discrimination, to the array of fixed alleles and incubating at 98°C for 5 minutes followed by step-wise decrease in temperature to 78°C over several hours in a limited volume containing 0.2M NaCl held on the array under a coverslip.

(xvi) addition of S1 nuclease and partial digestion of all mis-match containing duplexes, under appropriate conditions.

(xvii) stringent washing of the array to remove strands that failed to hybridize to the array and digested duplexes. This may be achieved by washing with SSC of low molar concentration at high temperature.

(xviii) detection of those alleles that formed mis-match free duplexes with the array. This may be achieved by observation of fluorescence providing VNTR alleles fixed to the array are appropriately labelled.

Example 6

In this example the $(CA)_n/(GT)_n$ VNTRs are generated from genomic DNA fragmented by a single restriction enzyme. The amplified products are hybridized to the genomic 'template' DNA of a single individual manifesting a particular trait of interest to yield the VNTR alleles represented in that genomic 'template' DNA. By hybridization of the amplified VNTR alleles to a sample of the same genomic 'template' DNA followed by partial digestion of any resulting mis-match containing duplexes and incubation of the products with biotin-dUTP, the spurious products of amplification due to reaction contamination and miss-priming events are removed. The non biotin containing alleles forming mis-match free duplexes are selected and hybridized to an array of VNTR alleles of known identity and spatial separation in which all possible alleles are represented, or a subset thereof. Providing VNTR alleles fixed to the array are appropriately labelled the identity of the alleles forming mis-match free duplexes with the array can be established.

(i) restriction digestion of genomic DNA using Rsa1 buffered appropriately at 37°C for several hours to yield genomic fragments.

(ii) Annealing of the sense and antisense oligonucleotides that comprise the adaptor (a 48mer and 12mer respectively). This may be achieved by mixing equimolar quantities of each oligo in ligase buffer and heating to 50°C followed by gradual cooling to 10°C over 1 hour.

(iii) Ligating the adaptor to the genomic fragments by addition of the fragments to the buffered adaptor and incubating at 15°C over night in the presence of T4 DNA ligase.

(iv) Removal of the ligase buffer and incubation of the ligated fragments with dideoxynucleotide triphosphates (ddNTPs) in the presence of taq DNA polymerase appropriately buffered at 72°C for 1 hour. Complete termination of the recessed 3' ends of the adaptor can be assessed by PCR of an aliquot of the reaction with adaptor primer which should fail to yield any products of amplification.

(v) Purification of the ligated fragments from the ddNTPs and amplification of those ligated fragments containing a $(CA)_n/(GT)_n$ VNTR under appropriate conditions with either a $(CA)_n$ primer and the adaptor primer, or a $(GT)_n$ primer and the adaptor primer. Over amplification of products should be avoided to limit the opportunity for formation of concatamers. Using PCR a total of about 25 cycles may be appropriate, though NASBA or other methods may be employed to achieve the same goal.

(vi) The $(CA)_n$ primers and the $(GT)_n$ primers are destroyed by physical separation of molecules based on molecular weight, electrophoretic mobility, or by digestion of single stranded DNA using an appropriate enzymes such as Exonclease 1 appropriately buffered at 37°C.

(vii) hybridization of the $(CA)_n$ primed products to an equimolar quantity of the genomic 'template' DNA of the single individual under investigation manifesting the particular trait of interest. Hybridization

should proceed under stringent conditions of high temperature and low salt concentration. Strand dissociation at 98°C for 5 minutes followed by a step-wise decrease in temperature to 78°C over several hours in solution of limited volume containing 0.2M NaCl may be appropriate.

(viii) incubation of the hybridization components with taq DNA polymerase appropriately buffered at 72°C to achieve extension of annealed strands.

(ix) addition of an amount of $(GT)_n$ primed products, similar to that of the $(CA)_n$ primed products, and adaptor primer to the reaction followed by amplification of products to yield VNTR alleles faithfully reproduced from the genomic template of the individual manifesting the trait of interest. This may be performed by PCR, NASBA or other methods.

(x) addition of the VNTR alleles to a molar excess of the same genomic 'template' DNA from which the alleles were derived followed by dissociation of all strands by incubation at 98°C for 5 minutes then step-wise reduction in temperature to 78°C over several hours in a limited reaction volume containing 0.2M NaCl. This allows the alleles to hybridize to the appropriate locus on the genomic DNA. Heteroduplexes that may occur result from miss-priming events or reaction contamination.

(xi) partial digestion of all mis-match containing duplexes using S1 nuclease under appropriate conditions.

(xii) incubation of the reaction components with biotin-dUTP in the presence of Taq DNA polymerase appropriately buffered at 72°C for 1 hour.

(xiii) separation of the biotin-containing molecules by binding to streptavidin on a solid support.

(xiv) repeated hybridization of the VNTR alleles to the same genomic 'template' DNA followed by partial digestion of mis-match containing duplex molecules and incubation with biotin-dUTP for removal of all spurious products of amplification by binding to streptavidin on a solid

support.

(xv) hybridization of the selected alleles with an array of all possible VNTR alleles that are fixed to a solid support with known identity and spatial separation This may be performed by addition of the selected VNTR alleles to the array of fixed alleles and incubating at 98°C for 5 minutes followed by step-wise decrease in temperature to 78°C over several hours in a limited volume containing 0.2M NaCl held on the array under a coverslip.

(xvi) addition of S1 nuclease and partial digestion of all mis-match containing duplexes, under appropriate conditions.

(xvii) stringent washing of the array to remove strands that failed to hybridize to the array and digested duplexes. This may be achieved by washing with SSC of low molar concentration at high temperature.

(xviii) detection of those alleles that formed mis-match free duplexes with the array. This may be achieved by observation of fluorescence providing the VNTR alleles fixed to the array are appropriately labelled.

Example 7

In this example the $(CA)_n/(GT)_n$ VNTRs are generated from genomic DNA fragmented by a single restriction enzyme. The amplified products are hybridized to the genomic 'template' DNA of a single individual manifesting a particular trait of interest to yield the VNTR alleles represented in that genomic 'template' DNA. By strand dissociation and reannealing of the amplified VNTR alleles followed by partial digestion of any resulting mis-match containing duplexes and incubation of the products with biotin-dUTP, the spurious products of amplification due to reaction contamination and miss-priming events are removed. The non-biotin containing duplexes are selected and are hybridized to an array of all possible VNTR alleles, or subset thereof, that are fixed to a solid support of

known identity and spatial separation. Mis-match containing duplexes formed in hybridization to the array are partially digested. Providing VNTR alleles fixed to the array are appropriately labelled the identity of the alleles forming mis-match free duplexes with the array can be established;

(i) restriction digestion of genomic DNA using Rsa1 buffered appropriately at 37°C for several hours to yield genomic fragments.

(ii) Annealing of the sense and antisense oligonucleotides that comprise the adaptor (a 48mer and 12mer respectively). This may be achieved by mixing equimolar quantities of each oligo in ligase buffer and heating to 50°C followed by gradual cooling to 10°C over 1 hour.

(iii) Ligating the adaptor to the genomic fragments by addition of the fragments to the buffered adaptor and incubating at 15°C over night in the presence of T4 DNA ligase.

(iv) Removal of the ligase buffer and incubation of the ligated fragments with dideoxynucleotide triphosphates (ddNTPs) in the presence of taq DNA polymerase appropriately buffered at 72°C for 1 hour. Complete termination of the recessed 3' ends of the adaptor can be assessed by PCR of an aliquot of the reaction with adaptor primer which should fail to yield any products of amplification.

(v) Purification of the ligated fragments from the ddNTPs and amplification of those ligated fragments containing a $(CA)_n/(GT)_n$ VNTR under appropriate conditions with either a $(CA)_n$ primer and the adaptor primer, or a $(GT)_n$ primer and the adaptor primer. Over amplification of products should be avoided to limit the opportunity for formation of concatamers. Using PCR a total of about 25 cycles may be appropriate, though NASBA or other methods may be employed to achieve the same goal.

(vi) The $(CA)_n$ primers and the $(GT)_n$ primers are destroyed by physical separation of molecules based on molecular weight, electrophoretic mobility, or by digestion of single stranded DNA using an

appropriate enzymes such as Exonclease 1 appropriately buffered at 37°C.

(vii) hybridization of the $(CA)_n$ primed products to an equimolar quantity of the genomic 'template' DNA of the single individual under investigation manifesting the particular trait of interest. Hybridization should proceed under stringent conditions of high temperature and low salt concentration. Strand dissociation at 98°C for 5 minutes followed by a step-wise decrease in temperature to 78°C over several hours in solution of limited volume containing 0.2M NaCl may be appropriate.

(viii) incubation of the hybridization components with taq DNA polymerase appropriately buffered at 72°C to achieve extension of annealed strands.

(ix) addition of an amount of $(GT)_n$ primed products, similar to that of the $(CA)_n$ primed products, and adaptor primer to the reaction followed by amplification of products to yield VNTR alleles faithfully reproduced from the genomic template of the individual manifesting the trait of interest. This may be performed by PCR, NASBA or other methods.

(x) strand dissociation and reannealing of the amplified VNTR alleles may be achieved by incubation at 98°C for 5 minutes then step-wise reduction in temperature to 78°C over several hours in a limited reaction volume containing 0.2M NaCl. Mis-match containing duplexes that may occur result from miss-priming events or reaction contamination.

(xi) partial digestion of the mis-match containing duplexes using S1 nuclease under appropriate conditions.

(xii) incubation of the reaction components with biotin-dUTP in the presence of Taq DNA polymerase appropriately buffered at 72°C for 1 hour.

(xiii) separation of the biotin-containing molecules by binding to streptavidin on a solid support.

(xiv) repeated strand dissociation, reannealing, partial digestion and incubation with biotin-dUTP for removal of all spurious products of

amplification by binding of mis-matched duplexes to streptavidin on a solid support.

(xv) hybridization of the selected mis-match free duplex forming alleles with an array of all possible VNTR alleles that are fixed to a solid support with known spatial separation This may be performed by addition of the selected VNTR alleles that have been selected by mis-match discrimination to the array of fixed alleles and incubating at 98°C for 5 minutes followed by step-wise decrease in temperature to 78°C over several hours in a limited volume containing 0.2M NaCl held on the array under a coverslip.

(xvi) addition of S1 nuclease and partial digestion of all mis-match containing duplexes, under appropriate conditions.

(xvii) stringent washing of the array to remove strands that failed to hybridize to the array and digested duplexes. This may be achieved by washing with SSC of low molar concentration at high temperature.

(xviii) detection of those alleles that formed mis-match free duplexes with the array. This may be achieved by observation of fluorescence providing VNTR alleles fixed to the array are appropriately labelled.

Example 8

Experimental Data

1. Preparation of amplimers using $(CA)_{13}$ and $(GU)_{13}$ primers.

(i) 2µg DNA was completely digested with 3µl *Rsa* I (10u/µl; Promega) in a total volume of 100µl:

8.5µl genomic DNA (equivalent to 3µg DNA)

10µl 10X reaction buffer

3µl *Rsa* I (10u/µl; Promega)

78.5µl dH_2O

100µl

The reaction was incubated at 37°C over night.

The enzyme was heat inactivated by incubation at 70°C for 20 minutes.

The DNA was separated from the buffer using a microconcentrator (Microcon-100; Amicon). The volume of DNA recovered = 10µl.

(ii) 2nmoles 48mer and 2nmoles 12mer oligonucleotides that constitute the adaptor were combined;

15.9µl 48mer (equivalent to 2 nmoles)

13.7µl 12mer (equivalent to 2 nmoles)

10µl 10X ligase buffer (NEB)

48.4µl dH_2O

88µl

The mixture was heated to 50°C and allowed to cool to 10°C over 1 hour.

To the 88µl of annealed adaptor was added the 10µl of digested DNA and ligation of the adaptor to the genomic fragments was performed:

88µl annealed adaptor/ ligase buffer (containing ATP)

10µl DNA

2µl T4 DNA ligase (400 NEBu/µl)

100µl

The reaction was incubated at 16°C over night.

The enzyme was heat inactivated by incubation at 70°C for 20 minutes.

The adaptor-ligated DNA fragments were separated from the buffer and non-ligated adaptor using a microconcentrator (Microcon-100; Amicon). The volume of DNA recovered = 12µl.

(iii) The adaptor-ligated DNA fragments were incubated with *Taq* DNA polymerase in the presence of dideoxynucleotide triphosphates to prevent 3' extension of the adaptor and non-ligated DNA in subsequent manipulations:

12µl microconcentrated DNA
3µl 10X NH4 reaction buffer
1µl 50mM $MgCl_2$
1µl ddATP
1µl ddCTP
1µl ddGTP
1µl ddTTP
1µl *Taq* DNA polymerase (5u/µl; Bioline)
9µl dH_2O
30µl

The reaction was incubated at 72°C for 2 hours.

The adaptor-ligated DNA with terminated 3' ends was purified by phenol/chloroform extraction and microconcentration.

The volume recovered was made up to 40µl and the concentration of DNA was gauged by gel quantification. The concentration of 75ng/µl was determined.

(iv) (CA) primed amplimers and (GU) primed amplimers were generated in separate reactions:

10µl 10X NH_4 reaction buffer
8µl 50mM $MgCl_2$
1.5µl 10mM dNTPs
1µl adaptor-ligated DNA with terminated 3' ends
4µl (CA) or (GU) primer (25pmol/µl)
73.5µl dH_2O
98µl

The reaction was overlaid with mineral oil.

The reaction was heated to 95°C for 2 minutes, during which time 1µl *Taq* DNA polymerase (5u/µl; Bioline) and 2µl adaptor primer (50pmol/µl) were added.

Thermal cycling was performed as follows:

95°C for 30 seconds, then 72°C for 45 seconds for a total of 20 cycles, followed by 72°C for 5 minutes.

To the 100µl of (CA) primed products was added 5µl Exonuclease I (10u/µl; Amersham Life Sciences) to remove the remaining (CA) primer. This reaction was incubated at 37°C for 30 minutes.

To the 100µl of (GU) primed products was added 10µl uracil-DNA glycosylase (1u/µl; NEB) to digest all uracil incorporated into the PCR products. This reaction was incubated at 37°C for 2 hours. 1µl 10mM dNTPs was added followed by 2µl T4 DNA polymerase (5u/µl; Epicentre laboratories) to remove the protruding (CA) strand that complemented the digested (GU) sequence. This reaction was incubated at 37°C for 5 minutes.

Both the pools of amplimers were phenol/chloroform extracted and microconcentrated (Microcon-100; Amicon). For each pool, the volume recovered were made up to 500µl, of which 5µl was analysed by spectrophotometry to determine the concentration of DNA.

2. Generation of VNTRs from genomic 'template' DNA

Equal amounts of (CA) and (GU) primed amplimers were hybridized to genomic 'template' DNA of a single individual prior to thermal cycling. In order to gauge the optimal ratio of amplimer to genomic 'template' DNA several reactions were performed using various amounts of 'template' DNA while keeping the amount of amplimers constant:

'Template' DNA (ng)	0	0.1	1	10	100	1000
Combined amplimers (ng)	1	1	1	1	1	1
5M NaCl (µl)	0.22	0.22	0.22	0.22	0.22	0.22
dH_2O (µl)	To a final volume of 5.55µl					

Each reaction was overlaid with mineral oil.

Each reaction was incubated at 98°C for 5 minutes, after which the temperature was reduced stepwise to 78°C over 4 hours.

The following was added to each hybridization:

5µl 10X NH_4 reaction buffer

4µl 50mM $MgCl_2$

0.75µl 10mM dNTPs

0.5µl adaptor primer (50pmol/µl)

34.2µl dH_2O

Each reaction was spun briefly in a microfuge.

Each reaction was heated to 72°C for 2 minutes and 0.5µl *Taq* DNA polymerase (5u/µl;Bioline) was added. The reactions were incubated at 72°C for a further 10 minutes, after which the temperature was raised to 95°C for 2 minutes. Thermal cycling was performed as follows;

95°C for 30 seconds, then 72°C for 1 minute, for a total of 10 cycles.

For each reaction 10µl of products amplified for 10 cycles were added to 40µl of reaction mix and amplified under the same conditions for an additional 22 cycles. 5µl of the ends products of amplification were run on an agarose gel. The reaction containing 100ng genomic 'template' DNA was found to yield the most products of amplification, equivalent to a ratio of 100:1 by mass of genomic 'template' DNA: amplimer.

The invention was validated by cloning the products of

amplification. Two colonies of *E.coli* that had successfully transformed were cultured, from which plasmids were later harvested. These plasmids were sequenced and were found to contain VNTR sequences at the multiple cloning sites.

Relevant Documents and References

1. Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S and Rafalski A (1996) The Comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* **2**, 225-238.

2. Karp A, Seberg O and Buiatti M (1996) Molecular Techniques in the Assessment of Botanical Diversity. *Annals of Botany* **78**, 143-149.

3. Lu J, Knox M R, Ambrose M J and Brown J K M (1996) Comparative analysis of genetic diversity in pea assessed by RFLP- and PRC-based methods. *Theoretical and Applied Genetics* **93**, 1103-1111.

4. Epplen J T, Buitkamp J, Bocker T and Epplen C (1995) Indirect gene diagnoses for complex (multifactorial) disease- a review. *Gene* **159**, 49-55.

5. Lisitsyn N A (1995) Representational difference analysis: finding the differences between genomes. *Trends in Genetics*, **11**, 303-307.

6. Mackill DJ, Zhang Z, Redona E D and Colowit P M (1996) Level of polymorphism and genetic mapping of AFLP markers in rice. *Genome* **39**, 969-977.

8. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M and Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407-4414.

9. EP, A, 0534858; Zabeau M *et al.*

10. WO 96/22388; Kuiper M *et al.*

CLAIMS

1. A method of making a mixture of VNTR alleles and their flanking regions of the genomic DNA of one or more members of a species of interest, which method comprises the steps of:

a) dividing genomic DNA of the species of interest into fragments,

b) ligating to each end of each fragment an adaptor having a recessed 3'-end which is blocked to prevent enzymatic chain extension, thereby forming a mixture of adaptor-terminated fragments,

c) using a portion of the mixture of adaptor-terminated fragments as templates with an adaptor primer and a VNTR primer to create a mixture of 5'-flanking VNTR amplimers,

d) using a portion of the mixture of adaptor-terminated fragments as templates with an adaptor primer and a VNTR antisense primer to create a mixture of 3'-flanking VNTR amplimers,

e) and using genomic DNA of the one or more members of the species of interest as template with the mixture of 5'-flanking VNTR amplimers and the mixture of 3'-flanking VNTR amplimers as primers to make the desired mixture of VNTR alleles and their flanking regions.

2. A method as claimed in claim 1 wherein the strands of the mixture of VNTR alleles and their flanking regions are separated and then re-annealed and any mis-matches are separated and discarded.

3. A method as claimed in claim 1 or claim 2, wherein there is used in step e) genomic DNA of one or more members of the species of interest which manifest a trait of interest, whereby the resulting mixture of VNTR alleles and their flanking sequences is representative of those which manifest the trait of interest.

4. A method as claimed in claim 3, wherein the mixture of VNTR alleles and their flanking sequences representative of those which manifest the trait of interest is hybridized with a mixture of VNTR alleles and their flanking sequences representative of those which do not manifest the trait of interest, and mis-matches are selected to provide a mixture of VNTR alleles and their flanking sequences which are characteristic of the trait of interest.

5. A portion of genomic DNA of one or more members of a species of interest, said portion consisting essentially of a representative mixture of alleles of a chosen VNTR sequence and their flanking regions.

6. A portion as claimed in claim 5, wherein the mixture of alleles is representative of those which exhibit a trait of interest.

7. A portion as claimed in claim 5 or claim 6, wherein each member of the mixture has an adaptor at its 3'-end and its 5'-end.

8. A portion of genomic DNA of a species of interest, said portion consisting essentially of a representative mixture of 3'-flanking regions of a chosen VNTR sequence, each member of the mixture carrying an adaptor at its 3'-end.

9. A portion of genomic DNA of a species of interest, said portion consisting essentially of a representative mixture of 5'-flanking regions of a chosen VNTR sequence, each member of the mixture carrying an adaptor at its 5'-end.

10. A method of treating a mixture of polymorphic alleles, the mixture being representative of those which show a trait of interest, which method comprises separating and then re-annealing strands of the mixture, and separating and discarding any mis-matches.

11. A method as claimed in claim 10, wherein the mixture of polymorphic alleles is a mixture of alleles of a chosen VNTR sequence and their flanking regions.

12. A method as claimed in claim 10 or claim 11, comprising the additional step of hybridizing the said mixture with a mixture of corresponding polymorphic alleles which are representative of those which do not show the trait of interest, and selecting mis-matches to provide a mixture of polymorphic alleles which are characteristic of the trait of interest.

13. A method as claimed in claims 11 and 12, wherein the mixture of corresponding polymorphic alleles is a mixture of alleles of the chosen VNTR sequence and their flanking regions, whereby there is obtained a mixture of VNTR alleles and their flanking sequences which are characteristic of the trait of interest.

14. A method of making a mixture of amplimers which method comprises the steps of:

a) dividing genomic DNA of one or more members of a species of interest into fragments,

b) ligating to each end of each fragment an adaptor having a recessed 3'-end which is blocked to prevent enzymatic chain extension, thereby forming a mixture of adaptor-terminated fragments, and

c) using a portion of the mixture of adaptor-terminated fragments as templates with an adaptor primer and a VNTR primer to create a mixture of 5'-flanking VNTR amplimers, and/or

d) using a portion of the mixture of adaptor-terminated fragments as templates with an adaptor primer and a VNTR antisense primer to create a mixture of 3'-flanking VNTR amplimers.

15. A kit comprising protocols and reagents for performing the method of any one of claims 1 to 14.

ABSTRACT

EXTRACTION AND UTILISATION OF VNTR ALLELES

The invention presented is a novel method for the extraction of VNTR alleles and for the concomitant detection of polymorphic markers for inherited traits at multiple loci by simultaneous comparison of complex genomes from multiple individuals. The product is designated a Total Representation of Alleles that are Informative for a Trait (TRAIT). These alleles may be used directly as genetic markers or may be used as vehicles to facilitate precise localisation of sequence variations responsible for a hereditary trait all organisms.